(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/064835 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US02/03168

(22) International Filing Date: 31 January 2002 (31.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/265,693 31 January 2001 (31.01.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/265,693 (CON)
Filed on 31 January 2001 (31.01.2001)

(71) Applicant *(for all designated States except US)*: **AMBION, INC.** [US/US]; 2130 Woodward Street, Suite 200, Austin, TX 78744 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BROWN, David** [US/US]; 1705 Drake, Austin, TX 78704 (US). **WINKLER, Matthew, M.** [US/US]; 960 Live Oak Circle, Austin, TX 78746 (US). **LAWRENCE, Frank** [US/US]; 2130 Woodward Street, Suite 200, Austin, TX 78744 (US). **SHELTON, Jeffrey** [US/US]; 2130 Woodward Street, Suite 200, Austin, TX 78744 (US).

(74) Agent: **WILSON, Mark, B.**; Fullbright & Jaworski L.L.P., Suite 2400, 600 Congress Avenue, Austin, TX 78701 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/064835 A2

(54) Title: METHODS FOR NUCLEIC ACID FINGERPRINT ANALYSIS

(57) Abstract: Methods are provided for generating nucleic acid fingerprints from complex nucleic acid populations. The methods rely on the addition of a nucleic acid tag with at least two functional domains to members of the complex population being assessed. Primers specific to the appended tag and arbitrary or adapter-specific primers are used to amplify subsets of the complex population. The amplified population is then used to generate labeled products for comparative expression analysis using the functional domain of the appended tag sequence that was not used for amplification. The separation of amplification and labeling substantially reduces the number of false positives common in amplification fingerprinting experiments by removing all amplification products from the analysis that do not derive from the anchored ends of the nucleic acid population.

DESCRIPTION

METHODS FOR NUCLEIC ACID FINGERPRINT ANALYSIS

BACKGROUND OF THE INVENTION

This patent application claims priority to U. S. Provisional Patent Application No. 60/265,693.

The present application was filed concurrently with: PCT Application No. _____ on January 31, 2002, entitled "COMPARATIVE ANALYSIS OF NUCLEIC ACIDS USING POPULATION TAGGING," which claims priority to U.S. Provisional Patent Application No. 60/265,694, filed on January 31, 2001, PCT Application No. _____ filed January 31, 2002, entitled "COMPETITIVE POPULATION NORMALIZATION FOR COMPARATIVE ANALYSIS OF NUCLEIC ACID SAMPLES," which claims priority to U. S. Provisional Patent Application No. 60/265,695 filed on January 31, 2001; and PCT Application No. _____, filed January 31, 2002 entitled "COMPETITIVE AMPLIFICATION OF FRACTIONATED TARGETS FROM MULTIPLE NUCLEIC ACID SAMPLES," which claims priority to U.S. Provisional Patent Application No. 60/265,692, filed on January 31, 2001. The disclosure of each of the above-identified applications is specifically incorporated herein by reference in its entirety without disclaimer.

1. Field of the Invention

The present invention relates generally to the fields of nucleic acid amplification. More particularly, it concerns using nucleic acid amplification to generate a nucleic acid fingerprint to define a complex nucleic acid population. The present invention also relates to methods for adding a nucleic acid tag sequence to a nucleic acid population to promote amplification and labeling of one or more subsets of a nucleic acid population.

2. Description of Related Art

A number of molecular biology methods have been developed to amplify nucleic acids of unknown sequence. Many of these procedures have employed addition of defined sequences to the ends of nucleic acid fragments prior to amplification. The additional sequences or "tags" are routinely used to create a

primer binding site or transcriptional promoter at the end of a nucleic acid population that is being sequenced or otherwise characterized.

U.S. Patent 5,104,792 describes a modification of PCR™ that allows amplification of nucleic acid fragments possessing unknown sequence. The primers for the amplification reaction contain random degenerate sequences at their 3' ends and a defined sequence at their 5' ends. Extension of the primers generates fragments containing unknown sequences that are flanked by the defined sequence. These fragments are amplified in a conventional PCR™ using primers that hybridize to the known flanking sequence. The method is used to amplify, clone, and characterize nucleic acid fragments that have not been sequenced.

Shuldiner *et al.*, (1990) describe a modification of RT-PCR in which a unique sequence is appended to the 5' end of cDNAs. Following reverse transcription, the cDNA is amplified using a primer directed to the unique sequence and a primer directed to a target-specific sequence. U.S. Patent 5,545,522 describes synthesizing cDNA from RNA using a complementary primer linked to an RNA polymerase promoter. After the cDNAs are converted to double strands, anti-sense RNA (aRNA) is transcribed from the cDNA using an RNA polymerase specific to the promoter region introduced during reverse transcription.

Co-amplification is commonly employed where target mRNAs are being quantified as it controls for the tube-to-tube variability of amplification efficiency that commonly affects product yield. Methods of co-amplification have been developed for PCR™ (U.S. Patent 5,213,961) and Nucleic Acid Sequence Based Amplification (NASBA) (U.S. Patents 5,834,255 and 5,837,501). Both of these procedures employ an exogenous standard that shares primer binding sites with the target sequence but whose amplification product is distinguishable from that of the target. The most common exogenous standards are RNA or DNA constructs whose concentration is known and whose amplification product is slightly smaller or larger than that of the target. Following amplification, the products derived from the target and the standard are distinguished by gel electrophoresis or hybridization to oligonucleotides bound to a solid support.

A variation of the competitive amplification scheme described above provides a method for co-amplifying targets in two different populations by PCR™ (Kato 1997; European Patent Application No. 98302726). The method, called

Adaptor-Tagged Competitive-PCR (ATAC-PCR), involves generating double-stranded cDNA from multiple RNA populations. The cDNAs are digested with a restriction enzyme and unique adapters are appended to the cDNAs via ligation at the restriction site. The adapters share a primer binding site but differ in size or by the presence of a restriction site. The adapter-tagged cDNAs are mixed and amplified with a gene-specific primer and a primer specific to the shared adapter sequence present at the proximal ends of the cDNA populations. If the adapters used for tagging different cDNA populations were different sizes, then the products of amplification are directly assessed by gel electrophoresis. If the adapters from the populations differ by a restriction site, then the amplified population must be aliquoted into different restriction digestion reactions to create distinguishable products that can be assessed by gel electrophoresis. Because the products arising from the populations are different sizes, they can be fractionated and quantified.

However, ATAC-PCR does not differentiate between amplification products that result from primers that interact with the adapter and those that interact with internal domains. ATAC-PCR relies on the direct assessment of the amplified population. The technology is primarily aimed at assessing single genes of known sequences in each amplification reaction. The ATAC-PCR patent and subsequent uses of the technology (Matoba 2000) describe its use to quantify single known targets in multiple populations. ATAC-PCR also relies on creating amplification products that are different sizes for the two populations being compared, making them incompatible with the side-by-side analysis employed in "fingerprint" analysis.

Techniques have been developed that incorporate nucleic acid amplification under low stringency annealing conditions to simultaneously amplify a plurality of nucleic acid targets in a sample (Welsh 1990). The nucleic acid products that result from amplification are fractionated by gel electrophoresis to create a set of bands referred to as a "fingerprint." A number of techniques have been developed for generating and analyzing amplification fingerprints to identify RNA or DNA targets that are unique or present at dissimilar levels between different nucleic acid populations.

The first technology for generating amplification fingerprints from nucleic acid populations was the Arbitrarily-Primed Polymerase Chain Reaction (AP-PCR) method described in U.S. Patents 5,487,985 and 5,861,245. AP-PCR involves adding

one or more oligonucleotide primers of 10-50 nucleotides to a DNA or cDNA population. The mixture of the sample nucleic acid and primer(s) are subjected to at least one cycle of PCR™. Products from the first round of primer extension are further amplified with the primer(s) already present in the reaction or by addition of other primer(s). The final amplified products are typically evaluated via gel electrophoresis, which creates a fingerprint pattern that reflects the composition of the sample nucleic acid population. Alternatively, the amplified products can be assessed by hybridization to target specific oligonucleotides that are present in solution or appended to a solid support (Trengle, 1999). Identical reactions with two or more nucleic acid populations create fingerprints that can be compared to identify those DNA fragments that are unique or present at different concentrations in the populations. Improvements to the technology have primarily involved making the procedure easier to perform, such as by using fewer primers per reaction, or less prone to false positives by using primers that allow higher stringency amplification conditions (U.S. Patent 5,861,245).

The most popular technique employing arbitrarily primed PCR is Differential Display-Reverse Transcription-PCR (DD-RT-PCR), described in U.S. Patent 5,262,311. An anchored oligo-dT primer (a primer with a poly-dT sequence with at least one non-dT residue at its 3' end) is used to prime the reverse transcription from the polyA tails of eukaryotic RNAs. The resulting cDNA is amplified by PCR™ using the same anchored oligo-dT primer used for reverse transcription and one or more primers of 9-20 nucleotides possessing some arbitrary sequence(s). The amplified products are typically displayed by gel electrophoresis. Any of the amplified fragments may be further analyzed, by removal, cloning, and sequencing.

As with AP-PCR, the primary utility of DD-RT-PCR is comparing the amplified fragment pattern of two or more identically amplified populations to identify target sequences that are unique or more or less prevalent in one population over another (U.S. Patent 5,599,672). Several variations to the DD-RT-PCR procedure have been described. U.S. Patent 5,665,547 describes using lower concentrations of dNTPs in the amplification reaction to improve specificity and reduce false positives. U.S. Patent 5,580,726 describes using longer arbitrary primers (greater than 20 nucleotides) with low stringency amplification conditions in the first few cycles and high stringency cycling for subsequent amplification to reduce the rate

of false positives. U.S. Patent 5,871,697 describes coupling DD-RT-PCR with a sequencing gel and a sequence database to allow the identification of amplified fragments based on the migration of the amplified product and the sequence of the arbitrary primer.

A third method of amplification fingerprint analysis uses restriction enzymes to fragment cDNA samples into small enough subsets to be analyzed. U.S. Patents 5,712,126 and 6,010,850 describe priming reverse transcription of an mRNA population with anchored oligo-dT. The first cDNA strand is converted to double strands by partially digesting the template RNA and priming second strand synthesis off of the resulting RNA primers. The double stranded material is digested by one or more restriction enzymes and one or more adapters are ligated at the restriction site(s). The modified population is amplified using PCR™ primers specific to the ligated adapter(s) and the reverse transcription primer. As with the techniques described above, the amplification products are fractionated by gel electrophoresis to create a pattern of bands that reflect the input RNA population. Multiple samples are typically processed under identical conditions and the resulting amplification products are compared to identify unique or differentially expressed RNAs.

Attempted improvements described for fingerprint amplification methods highlight the need to decrease the number of false positives that plague these techniques. For example, U.S. Patent 5,712,126 estimates that approximately 80% of the amplification products that appear to be differentially expressed in a DD-RT-PCR experiment turn out not to differ in relative expression level. In fact, when a single RNA sample from the same patient is split and the two resulting samples are taken through the DD-RT-PCR procedure, the fingerprint patterns typically differ by 5%. The difficulties with reproducibility are not surprising given the requirements of the amplification reactions. DD-RT-PCR requires that multiple amplification reactions generate identical amplification products except where differences in the populations exist.

However, nucleic acid amplification techniques are notorious for generating unexpected amplification products. An amplification reaction requiring primer annealing and extension from a complex nucleic acid mixture can create a variety of amplification products that are not predictable or consistent. These unpredictable or "spurious" amplification products can result from: multimerization of the arbitrary

primers, unpredictable interactions between the primers and the cDNA populations, or variable amplification between two sample populations. The inconsistency in generating fingerprints has kept the techniques from becoming a favored method for comparing RNA or DNA samples. Thus, there is still a need for better methods for arbitrarily amplifying members of a nucleic acid population for the purpose of identifying molecular differences between samples.

SUMMARY OF THE INVENTION

The present invention overcomes deficiencies in the art by the addition of bifunctional nucleic acid sequences or tags to target nucleic acids from one or more samples. These tags, when used in the methods of the present invention, allow the amplification and differentiation of target nucleic acid samples by the creation of one or more nucleic acid fingerprint populations in a manner that reduces the creation of false positives. In specific aspects, the present invention reduces the rate of false positives, typical among other comparative analysis techniques employing nucleic acid amplification, by removing those undesired nucleic acids from the fingerprint population that arise from spurious amplification.

In general, the invention relates to methods of analyzing one or more samples comprising:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a first nucleic acid tag, the nucleic acid tag comprising an amplification domain and a fingerprint domain;

c) amplifying the nucleic acid target using at least one adapter primer or arbitrary primer specific to a subset of the nucleic acid in the sample and one primer specific to the amplification domain, to produce at least a first amplified nucleic acid comprising the fingerprint domain and a segment of the first nucleic acid target;

d) generating labeled nucleic acid from the first amplified nucleic acid using the fingerprint domain; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

In some preferred embodiments the fingerprint domain of the first tag is appended between the at least a first nucleic acid target sequence and the amplification domain. In this manner, the fingerprint domain is assured of being amplified during the amplification process, and is present in the amplified nucleic acid. Of course, those of skill in the art will realize that there are other positions of the fingerprint and amplification domains in tags, and will be able to utilize tags with the domains in a variety of functional positions.

Further, while, at the most basic level, the methods of the invention may be employed with only one sample, in many cases, the first sample is only one of a plurality of samples. One of the advantages of the invention is the ability of it to be used to analyze many samples simultaneously. In preferred embodiments, the tags used for each sample will comprise a fingerprint domain that is unique to that sample. Of course, in cases where there are a plurality of samples, there will typically be a plurality of tags. Those of skill in the art will be able to employ the teachings of this specification to prepare appropriate tags. Typically, the number of unique tags required for a given procedure will be equal to the number of samples to be analyzed.

In some specific embodiments, the first nucleic acid tag is appended to a target nucleic acid in each of the plurality of samples to prepare a plurality of tagged nucleic acid samples and each tagged nucleic acid sample is amplified in a separate reaction.

In some specific embodiments, the method of the invention may be further defined as comprising:

a) obtaining at least a second sample;

b) preparing at least a second tagged nucleic acid sample by appending to at least the first nucleic acid target in the second sample at least a second nucleic acid tag, the second nucleic acid tag comprising the amplification domain and a second fingerprint domain, wherein the second fingerprint domain differs from the first fingerprint domain;

c) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a mixture of tagged nucleic acid samples;

d) amplifying the first nucleic acid target in the mixture, using at least one arbitrary or adapter primer and one primer specific to the amplification domain, to produce mixed amplified nucleic acid comprising both amplified nucleic acid from the first nucleic acid sample that comprises the first

fingerprint domain and a segment of the first nucleic acid target and amplified nucleic acid from the second nucleic acid sample that comprises the second fingerprint domain and a segment of the first nucleic acid target;

e) separating the mixed amplified nucleic acid into at least a first aliquot and a second aliquot;

f) generating first labeled nucleic acid from the first aliquot using the first fingerprint domain

g) generating second labeled nucleic acid from the second aliquot using the second fingerprint domain

h) fractionating the first labeled nucleic acid to create a first fingerprint; and

i) fractionating the second nucleic acid to create a second fingerprint.

While in even more specific embodiments, the invention my be further defined as comprising:

a) obtaining at least a third sample;

b) preparing at least a third tagged nucleic acid sample by appending to at least the first nucleic acid target in the third sample at least a third nucleic acid tag, the third nucleic acid tag comprising the amplification domain and a third fingerprint domain, wherein the third fingerprint domain differs from the first fingerprint domain and the second fingerprint domain;

c) mixing the first tagged nucleic acid sample, the second tagged nucleic acid sample, and the third tagged nucleic acid samples to create the mixture of tagged nucleic acid samples;

d) amplifying the first nucleic acid target in the mixture;

e) separating the mixed amplified nucleic acid into at least a first aliquot, a second aliquot, and a third aliquot;

f) generating first labeled nucleic acid, second labeled nucleic acid, and third labeled nucleic acid using the respective fingerprint domains; and

g) fractionating the first, second, and third, labeled nucleic acid to create respective first, second, and third fingerprints.

The methods may also be further defined as comprising:

a) obtaining a plurality of samples;

b) preparing a plurality of tagged nucleic acid samples by appending to at least the first nucleic acid target in each of the plurality of samples a plurality of nucleic acid tags, each nucleic acid tag comprising the amplification domain and a fingerprint domain, wherein the fingerprint domain for each sample is unique to that sample;

c) mixing the plurality of tagged nucleic acid samples to create a mixture of tagged nucleic acid samples;

d) amplifying the first nucleic acid target in the mixture, using at least one arbitrary primer or adapter primer and one primer specific to the amplification domain, to produce amplified nucleic acid comprising a plurality of amplified nucleic acids, each of which comprises a fingerprint domain and a segment of the first nucleic acid target;

e) separating the amplified nucleic acid into a plurality of aliquots;

f) generating labeled nucleic acid from each aliquot using a fingerprint domain unique to a different sample for each aliquot; and

g) fractionating the labeled nucleic acid from each aliquot to create a plurality of fingerprints.

In many applications, the nucleic acid target and/or the nucleic acid tag will be single-stranded nucleic acid. However this in not required in all embodiments of the invention, and those of skill will be able to follow the teachings of the specification to employ double-stranded nucleic acids in the invention. The nucleic acid target can be an RNA, DNA or a combination thereof. It is not required that the nucleic acid target be of natural origin, and the target can contain synthetic nucleotides. In specific aspects, the nucleic acid target is an RNA, for example, prokaryotic or eukaryotic RNA, total RNA, polyA RNA, an *in vitro* RNA transcript or a combination thereof. In other facets, the nucleic acid target may comprise DNA, such as, for example, cDNA , genomic DNA or a combination thereof. In certain aspects, at least one of the samples comprises nucleic acid isolated from a biological sample from, for example, a cell, tissue, organ or organism. In other aspects, at least one of the samples may comprise nucleic acid from an environmental sample. Of course, there is no need for all of the samples compared in a particular assay to be of the same source or type of source. A single sample may contain nucleic acid from a single source, or it may be the result of combining nucleic acids from multiple sources.

In some embodiments, amplifying comprises a PCR™ amplification reaction. In some specific cases, the primer specific to a subset of the nucleic acid in the sample is an arbitrary primer.

The amplification domain may comprise a primer binding domain, in which case amplifying comprises a primer annealing to and being extended from the primer binding domain.

In some embodiments, at least a first nucleic acid adapter is appended to the first nucleic acid target. For example, the first nucleic acid adapter can comprise a primer binding domain and amplifying can comprise a primer annealing to and being extended from the primer binding domain of the adapter.

Fractionating may comprise electrophoretic separation of the labeled DNA. Alternatively, fractionating may comprise array analysis of the labeled DNA, or any other method known to those of skill..

The fingerprint domain of the first nucleic acid tag or the second nucleic acid tag may comprise, for example, a primer binding domain, a transcription domain, an affinity domain, or a combination thereof. In some preferred embodiments, the fingerprint domain comprises at least one primer binding domain., and generating labeled nucleic acids comprises binding at least a first primer to at least one segment of the primer binding domain. Specific methods may comprise at least one primer extension reaction. In other embodiments, the fingerprint domain comprises at least one transcription domain, such as, for example, a promoter for a nucleic acid polymerase, and generating labeled nucleic acids comprises at least one transcription reaction. In some embodiments, the fingerprint domain comprises at least one affinity domain, and generating labeled nucleic acids comprises binding at least a first ligand to at least a segment of the affinity domain. In such cases, the first ligand may comprises a nucleic acid, protein, or other biomolecule, for example. The first ligand may bound to at least one solid support. Also, the first ligand may be labeled. The binding of the first ligand to the segment of the affinity domain can result in a detectable signal.

The first nucleic acid tag may further comprise at least one additional domain. The additional domain may be, for example, a labeling domain, a restriction enzyme domain, a secondary amplification domain, a secondary fingerprint domain or a combination thereof. In some preferred embodiments, the additional domain

comprises at least one labeling domain. In some cases, the labeling domain is comprised between the fingerprint domain and the amplification domain.

Some specific embodiments relate to methods of analyzing two or more samples comprising:

a) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of at least a first sample at least a first nucleic acid tag, the first nucleic acid tag comprising an amplification domain and a first fingerprint domain;

b) preparing at least a second tagged nucleic acid sample by appending to at least the first nucleic acid target in at least a second sample at least a second nucleic acid tag, the second nucleic acid tag comprising the amplification domain and a second fingerprint domain, wherein the second fingerprint domain differs from the first fingerprint domain;

c) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a mixture of tagged nucleic acid samples;

d) amplifying the first nucleic acid target in the mixture, using at least one arbitrary primer and one primer specific to the amplification domain, to produce amplified nucleic acid comprising both amplified nucleic acid from the first nucleic acid sample that comprises the first fingerprint domain and a segment of the first nucleic acid target and amplified nucleic acid from the second nucleic acid sample that comprises the second fingerprint domain and a segment of the first nucleic acid target in the second sample;

e) separating the amplified nucleic acid into at least a first aliquot and a second aliquot;

f) generating first labeled nucleic acid from the first aliquot using the first fingerprint domain

g) generating second labeled nucleic acid from the second aliquot using the second fingerprint domain

h) fractionating the first labeled nucleic acid to create a first fingerprint; and

i) fractionating the second nucleic acid to create a second fingerprint.

Additional embodiments relate to methods of analyzing one or more samples comprising:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a nucleic acid tag, the nucleic acid tag comprising an amplification domain and a fingerprint domain, wherein the fingerprint domain is appended between the nucleic acid target sequence and the amplification domain;

c) amplifying the nucleic acid target using at least one arbitrary primer and one primer specific to the amplification domain, to produce at least a first amplified nucleic acid comprising the fingerprint domain and a segment of the first nucleic acid target;

d) generating labeled nucleic acid from the first amplified nucleic acid using the fingerprint domain; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

While additional embodiment comprise methods of analyzing one or more samples comprising:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a first nucleic acid tag, the first nucleic acid tag comprising an amplification domain and a first fingerprint domain, wherein the fingerprint domain comprises a primer binding site and is appended between the nucleic acid target sequence and the amplification domain;

c) amplifying the nucleic acid target using at least one adapter primer or arbitrary primer and one primer specific to the amplification domain, to produce at least a first amplified nucleic acid comprising the first fingerprint domain and a segment of the first nucleic acid target;

d) generating labeled nucleic acid from the first amplified nucleic acid by primer extension from the primer binding site; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

Still more methods of analyzing one or more samples comprise:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a first nucleic acid tag, the first nucleic acid tag comprising an amplification domain and a first fingerprint domain, wherein the fingerprint domain comprises a transcription domain and is appended between the nucleic acid target sequence and the amplification domain;

c) amplifying the nucleic acid target using at least one adapter primer or arbitrary primer and one primer specific to the amplification domain, to produce at least a first amplified nucleic acid comprising the first fingerprint domain and a segment of the first nucleic acid target;

d) generating labeled nucleic acid from the first amplified nucleic acid using a nucleic acid polymerase specific to the transcription domain; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

The invention also relates to methods of analyzing one or more samples comprising:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a first nucleic acid tag, the first nucleic acid tag comprising an amplification domain and a first fingerprint domain, wherein the fingerprint domain comprises an affinity domain and is appended between the nucleic acid target sequence and the amplification domain;

c) amplifying the nucleic acid target using at least one adapter primer or arbitrary primer and one primer specific to the amplification domain, to produce at least a first amplified nucleic acid comprising the first fingerprint domain and a segment of the first nucleic acid target;

d) isolating or labeling nucleic acid from the first amplified nucleic acid using the affinity domain; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

As used herein, "plurality" means more than one. In certain specific aspects, a plurality may mean 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,

22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 250, 300, 400, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 7,500, 10,000, 15,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 125,000, 150,000, 200,000 or more, and any integer derivable therein, and any range derivable therein.

As used herein, "any integer derivable therein" means a integer between the numbers described in the specification, and "any range derivable therein" means any range selected from such numbers or integers.

As used herein the specification, "a" or "an" may mean one or more. As used herein in the claim(s), when used in conjunction with the word "comprising", the words "a" or "an" may mean one or more than one. As used herein "another" may mean at least a second or more.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1. Schematic shows reducing amplification artifacts using population tags.

FIG. 2. Schematic of tag composition.

FIG. 3. Schematic of tagged DD-RT-PCR.

FIG. 4. Schematic of Co-amplified, tagged DD-RT-PCR.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

A primary application of the invention is comparative expression analysis of nucleic acid populations or samples. In specific embodiments, the present invention

incorporates one or more amplification and labeling techniques to produce specific subsets of two or more nucleic acid populations. In particular aspects, these subsets are used to compare a plurality of nucleic acid targets in different nucleic acid samples. Thus, in particular aspects the invention provides a method of producing fingerprint populations that can be used to identify unique or quantitatively distinct sequences in RNA or DNA samples. In the context of this invention, a "fingerprint" or "fingerprint population" is defined as a set of detectable nucleic acid(s) that are produced during differentiation of an amplified nucleic acid population.

The present invention, in preferred embodiments, modifies members of a RNA or DNA population to share a common nucleic acid sequence called a tag. The process of appending tags to a population is referred to herein as "tagging." The techniques previously described, such as in U.S. Patents 5,104,792 and 5,545,522; European Patent Application #98302726; Jones and Winistorfer, 1992, and Shuldiner *et al.*, 1990, are typical of techniques for appending and using a nucleic acid sequence tag. However, none of the previous techniques considers using tags in conjunction with protocols, described herein, intended to generate unique fingerprints for one or more nucleic acid populations comprised in one or more samples.

The prior art further differs from the present invention in that the methods do not employ tandem functional sites that can be used to first amplify and subsequently create labeled nucleic acids for fingerprint analysis. None of the previously described methods considers using two or more adjacent tag domains to distinguish between amplification products that include the sequence derived from the reverse transcription primer and those amplification products that arise from unexpected and poorly reproducible interactions between the amplification primers and DNA in the amplification mixture. In contrast, other methods previously described for nucleic acid fingerprint analysis exclusively involve assessing the nucleic acids produced by amplification directly. In preferred embodiments, this is a key difference between other techniques and the present invention.

The invention is of particular use in procedures that simultaneously amplify many targets from a complex population in a manner that does not rely on prior knowledge of a target's sequence. For example, one method of the present invention is designed to identify differentially expressed genes (e.g., RNAs) in nucleic acid populations, whereas the other techniques use co-amplification of a target in a single

complex population and an added standard(s) (*e.g.*, a known sequence) to quantify the target(s). In contrast, a method of the present invention typically involves amplifying a plurality of targets of two or more complex populations. In particular embodiments, the present invention is aimed at assessing a plurality of targets of unknown sequence in multiple populations.

A. NUCLEIC ACIDS: TAGS AND SAMPLES

Embodiments of the present invention involve nucleic acids in many forms. Nucleic acid samples are collections of RNA and/or DNA derived or extracted from chemical or enzymatic reactions, biological samples, or environmental samples. Nucleic acid tags are nucleic acids of a defined sequence that are appended to nucleic acids in a sample to facilitate its analysis. There are many potential types of tags for use in the invention, which are described elsewhere in this specification.

1. General Description of Nucleic Acids

The general term "nucleic acid" is well known in the art. A "nucleic acid" as used herein will generally refer to a molecule (*i.e.*, a strand) of DNA, RNA or a derivative or analog thereof, comprising a nucleobase. A nucleobase includes, for example, a naturally occurring purine or pyrimidine base found in DNA (*e.g.*, an adenine "A," a guanine "G," a thymine "T" or a cytosine "C") or RNA (*e.g.*, an A, a G, an uracil "U" or a C). The term "nucleic acid" encompass the terms "oligonucleotide" and "polynucleotide," each as a subgenus of the term "nucleic acid." The term "oligonucleotide" refers to a molecule of between 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, and 100 nucleobases in length, and any range derivable therein. The term "polynucleotide" refers to at least one molecule of greater than about 100 nucleobases in length.

a. Nucleobases

As used herein a "nucleobase" refers to a heterocyclic base, such as for example a naturally occurring nucleobase (*i.e.*, an A, T, G, C or U) found in at least one naturally occurring nucleic acid (*i.e.*, DNA and RNA), and naturally or non-naturally occurring derivative(s) and analogs of such a nucleobase. A nucleobase

generally can form one or more hydrogen bonds ("anneal" or "hybridize") with at least one naturally occurring nucleobase in a manner that may substitute for naturally occurring nucleobase pairing (*e.g.*, the hydrogen bonding between A and T, G and C, and A and U).

"Purine" and/or "pyrimidine" nucleobase(s) encompass naturally occurring purine and/or pyrimidine nucleobases and also derivative(s) and analog(s) thereof, including but not limited to, a purine or pyrimidine substituted by one or more of an alkyl, caboxyalkyl, amino, hydroxyl, halogen (*i.e.*, fluoro, chloro, bromo, or iodo), thiol or alkylthiol moeity. Preferred alkyl (*e.g.*, alkyl, caboxyalkyl, etc.) moeities comprise of from about 1, about 2, about 3, about 4, about 5, to about 6 carbon atoms. Other non-limiting examples of a purine or pyrimidine include a deazapurine, a 2,6-diaminopurine, a 5-fluorouracil, a xanthine, a hypoxanthine, a 8-bromoguanine, a 8-chloroguanine, a bromothymine, a 8-aminoguanine, a 8-hydroxyguanine, a 8-methylguanine, a 8-thioguanine, an azaguanine, a 2-aminopurine, a 5-ethylcytosine, a 5-methylcyosine, a 5-bromouracil, a 5-ethyluracil, a 5-iodouracil, a 5-chlorouracil, a 5-propyluracil, a thiouracil, a 2-methyladenine, a methylthioadenine, a N,N-diemethyladenine, an azaadenines, a 8-bromoadenine, a 8-hydroxyadenine, a 6-hydroxyaminopurine, a 6-thiopurine, a 4-(6-aminohexyl/cytosine), and the like. A table non-limiting, purine and pyrimidine derivatives and analogs is also provided herein below.

Table 1-Purine and Pyrmidine Derivatives or Analogs			
Abbr.	**Modified base description**	**Abbr.**	**Modified base description**
Ac4c	4-acetylcytidine	Mam5s2u	5-methoxyaminomethyl-2-thiouridine
Chm5u	5-(carboxyhydroxylmethyl) uridine	Man q	Beta,D-mannosylqueosine
Cm	2′-O-methylcytidine	Mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
Cmnm5s2u	5-carboxymethylamino-methyl-2-thioridine	Mcm5u	5-methoxycarbonylmethyluridine
Cmnm5u	5-carboxymethylaminomethyluridine	Mo5u	5-methoxyuridine
D	Dihydrouridine	Ms2i6a	2-methylthio-N6-isopentenyladenosine
Fm	2'-O-methylpseudouridine	Ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
Gal q	Beta,D-galactosylqueosine	Mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine
Gm	2'-O-methylguanosine	Mv	Uridine-5-oxyacetic acid methylester
I	Inosine	O5u	Uridine-5-oxyacetic acid (v)
I6a	N6-isopentenyladenosine	Osyw	Wybutoxosine
M1a	1-methyladenosine	P	Pseudouridine
M1f	1-methylpseudouridine	Q	Queosine
M1g	1-methylguanosine	s2c	2-thiocytidine
M1I	1-methylinosine	s2t	5-methyl-2-thiouridine
M22g	2,2-dimethylguanosine	s2u	2-thiouridine
M2a	2-methyladenosine	s4u	4-thiouridine
M2g	2-methylguanosine	T	5-methyluridine
M3c	3-methylcytidine	t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine
M5c	5-methylcytidine	Tm	2'-O-methyl-5-methyluridine
M6a	N6-methyladenosine	Um	2'-O-methyluridine
M7g	7-methylguanosine	Yw	Wybutosine
Mam5u	5-methylaminomethyluridine	X	3-(3-amino-3-carboxypropyl)uridine, (acp3)u

A nucleobase may be comprised in a nucleside or nucleotide, using any chemical or natural synthesis method described herein or known to one of ordinary skill in the art.

b. Nucleosides

As used herein, a "nucleoside" refers to an individual chemical unit comprising a nucleobase covalently attached to a nucleobase linker moiety. A non-limiting example of a "nucleobase linker moiety" is a sugar comprising 5-carbon atoms (*i.e.*, a "5-carbon sugar"), including but not limited to a deoxyribose, a ribose, an arabinose, or a derivative or an analog of a 5-carbon sugar. Non-limiting examples of a derivative or an analog of a 5-carbon sugar include a 2'-fluoro-2'-deoxyribose or a carbocyclic sugar where a carbon is substituted for an oxygen atom in the sugar ring.

Different types of covalent attachment(s) of a nucleobase to a nucleobase linker moiety are known in the art. By way of non-limiting example, a nucleoside comprising a purine (*i.e.*, A or G) or a 7-deazapurine nucleobase typically covalently attaches the 9 position of a purine or a 7-deazapurine to the 1'-position of a 5-carbon sugar. In another non-limiting example, a nucleoside comprising a pyrimidine nucleobase (*i.e.*, C, T or U) typically covalently attaches a 1 position of a pyrimidine to a 1'-position of a 5-carbon sugar (Kornberg and Baker, 1992).

c. Nucleotides

As used herein, a "nucleotide" refers to a nucleoside further comprising a "backbone moiety". A backbone moiety generally covalently attaches a nucleotide to another molecule comprising a nucleotide, or to another nucleotide to form a nucleic acid. The "backbone moiety" in naturally occurring nucleotides typically comprises a phosphorus moiety, which is covalently attached to a 5-carbon sugar. The attachment of the backbone moiety typically occurs at either the 3'- or 5'-position of the 5-carbon sugar. However, other types of attachments are known in the art, particularly when a nucleotide comprises derivatives or analogs of a naturally occurring 5-carbon sugar or phosphorus moiety.

d. Nucleic Acid Analogs

A tag or other nucleic acid used in the invention may comprise, or be composed entirely of, a derivative or analog of a nucleobase, a nucleobase linker moiety and/or backbone moiety that may be present in a naturally occurring nucleic

acid. As used herein a "derivative" refers to a chemically modified or altered form of a naturally occurring molecule, while the terms "mimic" or "analog" refer to a molecule that may or may not structurally resemble a naturally occurring molecule or moiety, but possesses similar functions. As used herein, a "moiety" generally refers to a smaller chemical or molecular component of a larger chemical or molecular structure. Nucleobase, nucleoside and nucleotide analogs or derivatives are well known in the art, and have been described (see for example, Scheit, 1980, incorporated herein by reference).

Additional non-limiting examples of nucleosides, nucleotides or nucleic acids comprising 5-carbon sugar and/or backbone moiety derivatives or analogs, include those in U.S. Patent No. 5,681,947 which describes oligonucleotides comprising purine derivatives that form triple helixes with and/or prevent expression of dsDNA; U.S. Patents 5,652,099 and 5,763,167 which describe nucleic acids incorporating fluorescent analogs of nucleosides found in DNA or RNA, particularly for use as fluorescent nucleic acids probes; U.S. Patent 5,614,617 which describes oligonucleotide analogs with substitutions on pyrimidine rings that possess enhanced nuclease stability; U.S. Patents 5,670,663, 5,872,232 and 5,859,221 which describe oligonucleotide analogs with modified 5-carbon sugars (*i.e.*, modified 2'-deoxyfuranosyl moieties) used in nucleic acid detection; U.S. Patent 5,446,137 which describes oligonucleotides comprising at least one 5-carbon sugar moiety substituted at the 4' position with a substituent other than hydrogen that can be used in hybridization assays; U.S. Patent 5,886,165 which describes oligonucleotides with both deoxyribonucleotides with 3'-5' internucleotide linkages and ribonucleotides with 2'-5' internucleotide linkages; U.S. Patent 5,714,606 which describes a modified internucleotide linkage wherein a 3'-position oxygen of the internucleotide linkage is replaced by a carbon to enhance the nuclease resistance of nucleic acids; U.S. Patent 5,672,697 which describes oligonucleotides containing one or more 5' methylene phosphonate internucleotide linkages that enhance nuclease resistance; U.S. Patents 5,466,786 and 5,792,847 which describe the linkage of a substituent moeity which may comprise a drug or label to the 2' carbon of an oligonucleotide to provide enhanced nuclease stability; U.S. Patent 5,223,618 which describes oligonucleotide analogs with a 2 or 3 carbon backbone linkage attaching the 4' position and 3' position of adjacent 5-carbon sugar moiety to enhanced resistance to nucleases and

hybridization to target RNA; U.S. Patent 5,470,967 which describes oligonucleotides comprising at least one sulfamate or sulfamide internucleotide linkage that are useful as nucleic acid hybridization probe; U.S. Patents 5,378,825, 5,777,092, 5,623,070, 5,610,289 and 5,602,240 which describe oligonucleotides with three or four atom linker moeity replacing phosphodiester backbone moeity used for improved nuclease resistance; U.S. Patent 5,214,136 which describes olignucleotides conjugated to anthraquinone at the 5' terminus that possess enhanced hybridization to DNA or RNA; enhanced stability to nucleases; U.S. Patent 5,700,922 which describes PNA-DNA-PNA chimeras wherein the DNA comprises 2'-deoxy-erythro-pentofuranosyl nucleotides for enhanced nuclease resistance and binding affinity; and U.S. Patent 5,708,154 which describes RNA linked to a DNA to form a DNA-RNA hybrid.

e. Polyether and Peptide Nucleic Acids

In certain embodiments, it is contemplated that a tag or other nucleic acid comprising a derivative or analog of a nucleoside or nucleotide may be used in the methods and compositions of the invention. A non-limiting example is a "polyether nucleic acid", described in U.S. Patent Serial No. 5,908,845, incorporated herein by reference. In a polyether nucleic acid, one or more nucleobases are linked to chiral carbon atoms in a polyether backbone.

Another non-limiting example is a "peptide nucleic acid", also known as a "PNA", "peptide-based nucleic acid analog" or "PENAM", described in U.S. Patent Serial Nos. 5,786,461, 5891,625, 5,773,571, 5,766,855, 5,736,336, 5,719,262, 5,714,331, 5,539,082, and WO 92/20702, each of which is incorporated herein by reference. Peptide nucleic acids generally have enhanced sequence specificity, binding properties, and resistance to enzymatic degradation in comparison to molecules such as DNA and RNA (Egholm *et al.*, 1993; PCT/EP/01219). A peptide nucleic acid generally comprises one or more nucleotides or nucleosides that comprise a nucleobase moiety, a nucleobase linker moeity that is not a 5-carbon sugar, and/or a backbone moiety that is not a phosphate backbone moiety. Examples of nucleobase linker moieties described for PNAs include aza nitrogen atoms, amido and/or ureido tethers (see for example, U.S. Patent No. 5,539,082). Examples of backbone moieties described for PNAs include an aminoethylglycine, polyamide, polyethyl, polythioamide, polysulfinamide or polysulfonamide backbone moiety.

In certain embodiments, a nucleic acid analogue such as a peptide nucleic acid may be used to inhibit nucleic acid amplification, such as in PCR, to reduce false positives and discriminate between single base mutants, as described in U.S. Patent Serial No. 5891,625. Other modifications and uses of nucleic acid analogs are known in the art, and are encompassed by the invention. In a non-limiting example, U.S. Patent 5,786,461 describes PNAs with amino acid side chains attached to the PNA backbone to enhance solubility of the molecule. Another example is described in U.S. Patent Nos. 5,766,855, 5,719,262, 5,714,331 and 5,736,336, which describe PNAs comprising naturally and non-naturally occurring nucleobases and alkylamine side chains that provide improvements in sequence specificity, solubility and/or binding affinity relative to a naturally occurring nucleic acid.

f. Preparation of Nucleic Acids

A tag or other nucleic acid used in the invention may be made by any technique known to one of ordinary skill in the art, such as for example, chemical synthesis, enzymatic production or biological production. Non-limiting examples of a synthetic nucleic acid (*e.g.*, a synthetic oligonucleotide), include a nucleic acid made by *in vitro* chemically synthesis using phosphotriester, phosphite or phosphoramidite chemistry and solid phase techniques such as described in EP 266,032, incorporated herein by reference, or via deoxynucleoside H-phosphonate intermediates as described by Froehler *et al.*, 1986 and U.S. Patent Serial No. 5,705,629, each incorporated herein by reference. In the methods of the present invention, one or more oligonucleotides are used. Various different mechanisms of oligonucleotide synthesis have been disclosed in for example, U.S. Patents. 4,659,774, 4,816,571, 5,141,813, 5,264,566, 4,959,463, 5,428,148, 5,554,744, 5,574,146, 5,602,244, each of which is incorporated herein by reference.

A non-limiting example of an enzymatically produced nucleic acid includes one produced by enzymes in amplification reactions such as PCR (see for example, U.S. Patent 4,683,202 and U.S. Patent 4,682,195, each incorporated herein by reference), or the synthesis of an oligonucleotide described in U.S. Patent No. 5,645,897, incorporated herein by reference. A non-limiting example of a biologically produced nucleic acid includes a recombinant nucleic acid produced (*i.e.*, replicated) in a living cell, such as a recombinant DNA vector replicated in bacteria (see for example, Sambrook *et al.* 1989, incorporated herein by reference).

g. Nucleic Acid Purification

A tag or other nucleic acid used in the invention may be purified on polyacrylamide gels, cesium chloride centrifugation gradients, or by any other means known to one of ordinary skill in the art (see for example, Sambrook *et al.* 1989, incorporated herein by reference).

In particular embodiments, tags or other nucleic acid used in the invention may be isolated from at least one organelle, cell, tissue or organism. In certain embodiments, "isolated nucleic acid" refers to a nucleic acid that has been isolated free of, or is otherwise free of, the bulk of cellular components such as for example, macromolecules such as lipids or proteins, small biological molecules, and the like.

h. Nucleic Acid Complements

The present invention also encompasses a nucleic acid that is complementary to a specific nucleic acid sequence. A nucleic acid "complement(s)" or is "complementary" to another nucleic acid when it is capable of base-pairing with another nucleic acid according to the standard Watson-Crick, Hoogsteen or reverse Hoogsteen binding complementarily rules. As used herein "another nucleic acid" may refer to a separate molecule or a spatial separated sequence of the same molecule.

i. Hybridization

As used herein, "hybridization", "hybridizes" or "capable of hybridizing" is understood to mean the forming of a double or triple stranded molecule or a molecule with partial double or triple stranded nature. The term "anneal" as used herein is synonymous with "hybridize." The term "hybridization", "hybridize(s)" or "capable of hybridizing" encompasses the terms "stringent condition(s)" or "high stringency" and the terms "low stringency" or "low stringency condition(s)."

As used herein "stringent condition(s)" or "high stringency" are those conditions that allow hybridization between or within one or more nucleic acid strand(s) containing complementary sequence(s), but precludes hybridization of random sequences. Stringent conditions tolerate little, if any, mismatch between a nucleic acid and a target strand. Such conditions are well known to those of ordinary skill in the art, and are preferred for applications requiring high selectivity. Non-limiting applications include isolating a nucleic acid, such as a gene or a nucleic acid

segment thereof, or detecting at least one specific mRNA transcript or a nucleic acid segment thereof, and the like.

Stringent conditions may comprise low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about 50°C to about 70°C. It is understood that the temperature and ionic strength of a desired stringency are determined in part by the length of the particular nucleic acid(s), the length and nucleobase content of the target sequence(s), the charge composition of the nucleic acid(s), and to the presence or concentration of formamide, tetramethylammonium chloride or other solvent(s) in a hybridization mixture.

It is also understood that these ranges, compositions and conditions for hybridization are mentioned by way of non-limiting examples only, and that the desired stringency for a particular hybridization reaction is often determined empirically by comparison to one or more positive or negative controls. Depending on the application envisioned it is preferred to employ varying conditions of hybridization to achieve varying degrees of selectivity of a nucleic acid towards a target sequence. In a non-limiting example, identification or isolation of a related target nucleic acid that does not hybridize to a nucleic acid under stringent conditions may be achieved by hybridization at low temperature and/or high ionic strength. Such conditions are termed "low stringency" or "low stringency conditions", and non-limiting examples of low stringency include hybridization performed at about 0.15 M to about 0.9 M NaCl at a temperature range of about 20°C to about 50°C. Of course, it is within the skill of one in the art to further modify the low or high stringency conditions to suite a particular application.

B. NUCLEIC ACID SAMPLES (POPULATIONS)

The invention can be applied to the comparative analysis of any nucleic acid population. The nucleic acids can be RNA, DNA, or both. The nucleic acids can be part of a collection of other molecules, including proteins, carbohydrates or small molecules. While the population can comprise even a single sequence, the method is best suited for nucleic acid samples that include hundreds or thousands of unique sequences.

The terms "target", "target nucleic acid" and "target sequence" refer to one or more nucleic acids (*e.g.*, DNA, RNA) of a specific sequence that are being

characterized. Often, target nucleic acids comprise a sub-population of nucleic acids relative to all the nucleic acid sequences originally present in a nucleic acid sample.

1. Sources of Nucleic Acid Samples

Nucleic acid samples can be obtained from biological material, such as cells, tissues, organs or organisms. The invention is particularly relevant to total and polyA RNA preparations from tissues or cells. Similarly, the invention could be applied to cDNAs derived from cells or tissues. In other embodiments, multiple genomic DNA samples could be assessed using the methods of the present invention.

a. Cells and Tissues

A cell, or a tissue comprising cells, may be a source of nucleic acids for the present invention. In certain embodiments, cells or tissue may be part of or separated from an organism. In certain embodiments, a cell or tissue may comprise, but is not limited to, adipocytes, alveolar, ameloblasts, axon, basal cells, blood (*e.g.*, lymphocytes), blood vessel, bone, bone marrow, brain, breast, cartilage, cervix, colon, cornea, embryonic, endometrium, endothelial, epithelial, esophagus, facia, fibroblast, follicular, ganglion cells, glial cells, goblet cells, kidney, liver, lung, lymph node, muscle, neuron, ovaries, pancreas, peripheral blood, prostate, skin, skin, small intestine, spleen, stem cells, stomach, testes, anthers, ascites, cobs, ears, flowers, husks, kernels, leaves, meristematic cells, pollen, root tips, roots, silk, stalks, and all cancers thereof.

b. Organisms

In certain embodiments, an organism may be a source of nucleic acids for the present invention. In certain embodiments, the organism may be, but is not limited to, a eubacteria, an archaea, a eukaryote or a virus (for example, webpage http://phylogeny.arizona.edu/tree/phylogeny.html).

i. Eubacteria

In certain embodiments, the organism is a eubacteria. In particular embodiments, the eubacteria may be, but is not limited to, an aquifecales; a thermotogales; a thermodesulfobacterium; a member of the thermus-deinococcus group; a chloroflecales; a cyanobacteria; a firmicutes; a member of the leptospirillum group; a synergistes; a member of the chlorobium-flavobacteria group; a member of the chlamydia-verrucomicrobia group, including but not limited to a verrucomicrobia

or a chlamydia; a planctomycetales; a flexistipes; a member of the fibrobacter group; a spirochetes; a proteobacteria, including but not limited to an alpha proteobacteria, a beta proteobacteria, a delta & epsilon proteobacteria or a gamma proteobacteria. In certain aspects, an organelle derived from eubacteria are contemplated, including a mitochondria or a chloroplast.

ii. Archaea

In certain embodiments, the organism is an archaea (a.k.a. archaebacteria; *e.g.,* a methanogens, a halophiles, a sulfolobus). In particular embodiments, the archaea may be, but is not limited to, a korarchaeota; a crenarchaeota, including but not limited to, a thermofilum, a pyrobaculum, a thermoproteus, a sulfolobus, a metallosphaera, an acidianus, a thermodiscus, a igneococcus, a thermosphaera, a desulfurococcus, a staphylothermus, a pyrolobus, a hyperthermus or a pyrodictium; or an euryarchaeota, including but not limited to a halobacteriales, methanomicrobiales, a methanobacteriales, a methanococcales, a methanopyrales, an archeoglobales, a thermoplasmales or a thermococcales.

iii. Eukaryotes

In certain embodiments, the organism is a eukaryote (*e.g.,* a protist, a plant, a fungi, an animal). In particular embodiments, the eukaryote may be, but is not limited to, a microsporidia, a diplomonad, an oxymonad, a retortamonad, a parabasalid, a pelobiont, an entamoebae or a mitochondrial eukaryote (*e.g.,* an animal, a plant, a fungi, a stramenopiles).

iv. Viruses

In certain embodiments the organism may be a virus. In particular aspects, the virus may be, but is not limited to, a DNA virus, including but not limited to a ssDNA virus or a dsDNA virus; a DNA RNA rev transcribing virus; a RNA virus, including but not limited to a dsRNA virus, including but not limited to a -ve stranded ssRNA or a +ve stranded ssRNA; or an unassigned virus.

c. Synthetic Samples

Nucleic acid samples comprising populations designed by the hand of man may also be generated and used as a standard against which another sample or subpopulation of target sequences could be compared. For example, a synthetic sample may comprise a collection of nucleic acids (*e.g.,* RNA, cDNA or genomic

DNA) from many different tissues, cells (*e.g.*, cell cultures), or other samples that could provide an average population against which a sample, or subpopulation of target sequences, could be compared. In another non-limiting example, the synthetic sample could comprise a collection of *in vitro* transcripts sharing a specific tag sequence so that they could be co-amplified with nucleic acids from another sample (*e.g.*, RNA) to quantify a collection of targets. In another example, the synthetic sample could comprise a set of DNAs sharing a specific tag sequence that could be used to quantify a sample comprising a target DNA population.

d. Sample Mixtures

Embodiments employing co-amplification of two or more nucleic acid samples require that the different samples be distinguishable. This is achieved by using unique fingerprint domains for the different samples. Differentially tagged samples are mixed prior to amplification to create a sample mixture. Following amplification, fingerprints unique to each sample are generated using separate reactions specific to each fingerprint domain.

C. FUNCTIONAL CHARACTERISTICS OF TAGS

The invention involves appending a tag to one or more target sequences, up to all nucleic acid sequences, comprised in a nucleic acid population. A tag is a common sequence shared by various nucleic acid targets of a sample's nucleic acid population. In preferred embodiments, a tag is an RNA, DNA, or other molecule that can be used as a template by a polymerase to generate a complementary strand.

A tag preferably comprises at least two functional domains. The first, referred to as a "fingerprint domain", can be used to created labeled nucleic acid fingerprints from amplified nucleic acid samples. The second functional element, referred to as an "amplification domain," is used for amplifying a subset of a tagged population. Of course, a tag may comprise one or more additional sequences. Generally, additional sequences will possess functional properties, such as, for example, a property that facilitates analysis of tagged nucleic acids, amplified nucleic acids, and/or differentiated nucleic acids, particularly nucleic acids that comprise a unique fingerprint for a nucleic acid sample.

It is particularly preferred that the fingerprint domain be between the amplification domain and the sequence of each target nucleic acid in the sample. In

other words, it is particularly preferred that a fingerprint domain is internal to the amplification domain.

The fingerprint domain and amplification domain sequences can overlap, though it is particularly preferred that they are functionally distinct.

In certain aspects of the invention, the amplification domains for different populations are the same and the fingerprint domains are different. Differentially tagged samples may be mixed and amplified using a primer specific to the amplification domain of the tags and primer(s) possessing arbitrary or adapter-specific sequences. The unique fingerprint domains can then be used to generate labeled nucleic acid fingerprints specific to each of the samples.

1. Amplification Domains

It is particularly preferred that a tag comprise at least one amplification domain. As used herein, an amplification domain will primarily be a sequence that can support the amplification of a nucleic acid that comprises such a sequence. Use of nucleic acid sequences in amplification reactions are well known in the art, and non-limiting examples are described herein.

In certain embodiments, an amplification domain will comprise a sequence that can support primer binding and extension. Standard rules for primer design apply (Sambrook, 1989). In specific aspects, an amplification domain will preferably comprise a primer binding site for PCR™ amplification. PCR™ does not require any specialized structure or sequence to sustain amplification; the PCR™ amplification primer typically contains only binding sequences. Parameters for primer design for PCR™ are well known in the art (see, *e.g.*, Beasley *et al.*, 1999).

Primer binding sites for other types of amplification methods might also be used in amplification domains. Often such primer binding regions share similar characteristics with PCR™ primer binding sites, however the primers used for other amplification methods typically possess sequences 5' to the binding domain. For instance, primers for 3SR and NASBA contain an RNA polymerase promoter sequence 5′ to the priming site to support subsequent transcription. Because 3SR and NASBA are performed at relatively low temperature (37°C to 42°C), the amplification domains can have much lower melting temperatures than those used for PCR™.

In embodiments wherein a sample mixture is assessed, the amplification domains of the tags used in the samples that were mixed will preferably be identical to facilitate co-amplification of the target sequences.

2. Fingerprint domains

It is particularly preferred that a tag comprise at least one fingerprint domain. A fingerprint domain comprises a sequence that can be used to convert amplified nucleic acid(s) into DNA or RNA that can be readily analyzed. In some embodiments, the fingerprint domain can be used to identify the sample from which a particular amplified nucleic acid derives. For example, a fingerprint domain may comprise an affinity sequence for removing amplified nucleic acids from a sample, a primer binding site for labeled DNA synthesis, or a transcription promoter for labeled RNA synthesis. In most embodiments, it is preferred that a fingerprint domain be able to create a population of labeled nucleic acids for analysis.

In embodiments where two or more nucleic acid samples are mixed and co-amplified, fingerprint domains should be unique for each sample. Following amplification, labeling reactions specific to each fingerprint domain can be used to generate fingerprints of each sample.

a. Primer Binding Domains

A fingerprint domain may comprise a primer binding site (a "primer binding domain"). For example, a primer binding site may provide an annealing site for various types of primers that can be extended by a polymerase to generate a labeled nucleic acid (*e.g.*, DNA). Binding sites for primers are well known in the art (Sambrook 1989).

b. Transcription Domains

In certain embodiments, a fingerprint domain may comprise a promoter sequence (a "transcription domain") that binds an RNA polymerase to initiate transcription. Labeled RNA resulting from transcription can be used for fingerprint analysis. For example, an amplified population possessing promoter sequences can be transcribed in a reaction (*e.g.*, an *in vitro* reaction) with one or more labeled nucleotides (radio- or non-isotopic-labeled NTPs) and an appropriate RNA polymerase to convert amplification products into labeled RNA that can be used for fingerprint analysis.

c. **Affinity Domains**

In certain embodiments, a fingerprint domain may provide an affinity site for hybridization or binding (an "affinity domain") to a sample's amplified and/or labeled nucleic acid(s) to create a fingerprint. An affinity domain may bind or hybridize to a ligand comprising, but not limited to, a nucleic acid, protein or other molecule. For example, an affinity domain may be used to extract one or more labeled nucleic acids from an amplified population. In certain aspects, the affinity site may comprise a sequence that can hybridize to a nucleic acid ligand (*e.g.*, an oligonucleotide or polynucleotide). In cases involving co-amplification of sample mixtures, the nucleic acid ligand may provide a method for identifying a sample-specific tag sequence. For example, the sample specific ligands may be bound to different solid supports so that nucleic acids possessing a fingerprint domain with an affinity for that ligand can be readily separated and analyzed separate from a sample mixture.

3. Additional Functional Domains

A tag may comprise one or more additional functional or structural sequences, as described herein or as would be known to one of ordinary skill in the art. In certain embodiments, these domains may be partly or fully comprised within other domains, such as, for example an amplification domain or a fingerprint domain. In other embodiments, these additional domains may be comprised in sequences that do not comprise the amplification domain or fingerprint domain.

These additional domain(s) may be used to support additional molecular biological reactions, including but not limited to, a nested amplification reaction, a secondary differentiation reaction, a labeling reaction, a restriction digestion reaction, a cloning reaction, a hybridization reaction, sequencing reaction or a combination thereof. The addition of one or more additional domains will be particularly preferred in certain embodiments for manipulating one or more samples including samples that comprise a sample mixture.

Additional sequences described herein are by no means intended as an exhaustive list of all of the potential functional domains that can be included to facilitate production, amplification, fingerprint generation, comparison or analysis of a nucleic acid target and/or sample. The list is merely intended to provide examples of some requirements and benefits of additional functional domains that can be incorporated into the nucleic acid tag.

a. Labeling Domains

A tag may comprise a sequence that is used in a labeling reaction (a "labeling domain") to convert an amplified nucleic acid population into a labeled product population for subsequent fingerprint generation via affinity domains. A variety of sequences can be used to support the production of labeled products, and non-limiting examples are described herein. In specific embodiments, a labeling domain may be used for labeled DNA or labeled RNA product synthesis. It is particularly preferred that the labeling domain be situated upstream of the fingerprint domain so that the labeled population include the fingerprint domain sequence.

b. Sequencing Primer Binding Sites

A tag may comprise a primer binding site for a sequencing primer. For example, in certain preferred embodiments a primer binding site could be included in the tag sequence, either within the fingerprint domain or between the fingerprint domain and the target nucleic acid's sequence. Following fingerprint analysis, nucleic acid fingerprint products can be excised, cloned, and sequenced using the sequencing primer binding site.

c. Restriction Enzyme Sites

A tag sequence may comprise one or more selected restriction enzyme sites. In some embodiments, a restriction enzyme site may facilitate cloning of a nucleic acid comprising a tag. Methods of cloning are common in the art (Sambrook 1989). For example, cloning the nucleic acids from interesting fingerprint bands will be particularly preferred to facilitate sequence analysis or additional characterization.

d. Secondary Amplification Domains

One or more amplification domains in addition to the primary amplification domain may be used for nested amplification(U.S. Patent 5,340728). In general embodiments, nested amplification comprises sequential amplification reactions wherein a first amplification with a first set of one or more primers generates one or more primary amplified nucleic acids, and at least a second amplification of the one or more primary amplified nucleic acids with another set of primers comprising a primer that binds a sequence partly or fully internal to a primer of the first set, so that a nucleic acid segment of the one or more primary amplified nucleic acids is then amplified to produce one or more secondary amplified nucleic acids. In certain

embodiments, nested amplification might be required for those targets that are present in only a few copies in a sample or where small amounts of a sample (*e.g.*, a few mammalian cells) are available. In certain embodiments, a secondary amplification domain may be located between the primary amplification domain and the primary fingerprint domain.

e. Secondary Fingerprint Domains

One or more fingerprint domains in addition to the primary fingerprint domain may be used in conjunction with the primary fingerprint domain to further distinguish amplification products. For example, if transcription is being used to produce fingerprints of samples that are part of a sample mixture, and only a few different polymerases are available for *in vitro* transcription, then only a few samples can be assayed at a time. Incorporating a secondary fingerprint domain between the amplification domain and the primary fingerprint domain would allow additional samples to be mixed and assayed by the methods of the present invention.

Incorporating a primer binding site between the amplification domain and the primary fingerprint domain would allow additional sample populations to be mixed. In one aspect, several samples could use tags with the same transcription promoter that comprises their primary fingerprint domain so long as their secondary fingerprint domains were unique. The primary amplification would use a single tag-specific primer for all samples. The amplified population could then be split and further amplified with primers specific to the secondary fingerprint domains. Each of the resulting samples could then be used to generate labeled nucleic acids for analysis, using the different transcription promoters.

D. METHODS FOR APPENDING TAGS TO POPULATIONS

A nucleic acid tag of the present invention may be added to or appended to a nucleic acid population. As would be appreciated by one of ordinary skill in the art, different methods of tag attachment or incorporation may be used depending on whether the nucleic acid population comprises DNA or RNA. Non-limiting examples of such methods that may be used are described herein, though other methods can be used as would be known by one of ordinary skill in the art.

1. Tagging RNA

The methods of the present invention are applicable to tagging both eukaryotic RNA and/or prokaryotic RNA. In other aspects, the present invention may be applied to tag polyA selected or total RNA populations.

a. Ligation

In certain embodiments, a tag can be appended to the 3' ends of RNAs by a ligase (*e.g.*, an enzymatic protein, nucleic acid or chemical that induces ligation). For ligation, an excess of RNA or DNA polynucleotide tag possessing a 5' phosphate can be added to a RNA population. Incubation of the mixture with a ligating agent (*e.g.*, RNA ligase) generates RNAs with the tag ligated to the 3' end of the RNAs.

In general embodiments, more efficient ligation may be achieved by adding bridging oligonucleotides to the ligation reaction. Hybridization of a bridge to both a nucleic acid to be tagged (*e.g.*, an RNA or DNA in the sample) and a tag will align the 3' and 5' ends of the two molecules, enhancing ligation efficiency. In a non-limiting example, a bridging oligonucleotide may comprise a sequence at its 3' end that is complementary to the 3' end of the RNA population and a sequence at its 5' end that is complementary to the 5' end of the tag.

b. Cap Dependent Ligation

In one embodiment, a cap dependent ligation may be used to selectively append tags to the 5' ends of eukaryotic mRNAs. In general aspects, an RNA may be tagged by the combined enzymatic activities of a phosphatase, a pyrophosphatase (*e.g.*, a tobacco acid pyrophosphatase) that leaves a 5' phosphate at the 5' terminus of a capped message, and nucleic acid ligase (*e.g.*, an RNA ligase).

In a non-limiting example, a total RNA population is treated with a phosphatase, such as calf intestinal phophatase (CIP), to dephosphorylate (*i.e.*, remove the 5' terminal phosphate cap structure of euakaryotic mRNA) of the RNA population. CIP is specific to RNAs with free terminal phosphates, therefore the 5' phosphates of rRNAs, tRNAs, and partially degraded mRNAs are removed leaving these RNAs with 5' hydroxyls. After the CIP is inactivated, the RNA preparation is treated with a phosphatase such as tobacco acid pyrophoshatase (TAP) to convert the 5' cap structures of mRNAs to 5' monophosphates. An excess of a DNA or RNA polynucleotide tag is added to the RNA population as well as a ligase

that functions on RNA substrates. The tag should ligate exclusively to TAP modified RNAs possessing 5' monophosphates as all of the non-capped RNAs possess 5' hydroxyls following CIP treatment, only the mRNAs should be ligated to the tag. The resulting tagged mRNA population can be reverse transcribed, amplified and the amplified nucleic acids used to generate a fingerprint.

c. Enzymatic Polymerization

In an additional embodiment, a tag is incorporated into an RNA population by enzymatic polymerization. A tag comprising a 3' nucleotide that cannot be extended by polymerization (see for example, U.S. Patent 6,057,134), can be hybridized to the 3' ends of an RNA population. An RNA or DNA polymerase with the ability to extend primer template junctions can be added to the mixture and allowed to extend the 3' ends of the RNAs in the population, incorporating a sequence complementary to the hybridized oligonucleotide at the 3' ends of the RNA in the sample. If the nucleic acid that serves as a template comprises a tag sequence, then the polymerization reaction effectively tags the nucleic acid sample population.

d. CAPswitch™

A method for tagging mRNAs by Cap-induced primer extension is described in U.S. Patent 5,962,271. The technology, referred to as CAPswitch™, uses a unique CAPswitch oligonucleotide in the first strand cDNA synthesis reaction. When reverse transcriptase stops at the 5' end of an mRNA template in the course of first strand cDNA synthesis, it switches to a CAPswitch oligonucleotide and continues DNA synthesis to the end of a CAPswitch oligonucleotide. The resulting cDNA has at its 3' end a sequence that is complementary to the CAPswitch oligonucleotide sequence. The CAPswitch technology may be used to tag one or more RNA populations by using one or more CAPswitch oligonucleotides comprising fingerprint and amplification domains.

2. Tagging RNA Populations by Reverse Transcription

In a preferred embodiment, tag sequences may be appended to sample nucleic acids by reverse transcription. For example, tagged cDNA populations can be conveniently generated by priming reverse transcription with oligonucleotides tags comprising an amplification and an fingerprint domain at its 5' end and a priming domain at its 3' end. Hybridization of the primer to one or more species in an RNA

sample and subsequent reverse transcription yields cDNA with tag sequences at its 5' end.

For example, most eukaryotic mRNAs possess a polyA tail that can be tagged with a primer that has a polyT or polyU at or near its 3' end and an amplification and a fingerprint domain at its 5' end. The polyA specific tag primer can be extended from the polyA tail of the mRNAs. The resulting cDNAs would possess the tag sequences at or near their 5' ends that may be used in one or more amplification or fingerprint reactions.

3. Tagging Prokaryotic RNA Samples

The methods described above may be less preferred for tagging a non-polyA RNA, such as a prokaryotic RNA. However, analysis of prokaryotic RNA samples is desirable in certain aspects.

In a non-limiting example, it may be desirable to remove or separate small RNAs (*e.g.*, tRNAs) from prokaryotic mRNA or an RNA population lacking a polyA tail. Methods of removing small RNA are known to those of skill in the art, and include such methods as a lithium chloride precipitation. Lithium chloride precipitation is specific to RNAs greater than 100-300 nucleotides, thus tRNAs and other small RNAs will be removed from the RNA population (Sambrook 1989).

In embodiments wherein non-rRNAs are the target RNAs, one or more rRNA specific oligonucleotides and a polyA polymerase can be used to add polyA tails to a target RNA. For example, a sample comprising prokaryotic total RNA can be precipitated with lithium chloride. After removal of the solution comprising the tRNAs, a resulting RNA population can be resuspended and hybridized to one or more oligonucleotides complementary to the 3' ends of the major prokaryotic rRNAs. The 5' ends of the oligonucleotide(s) will preferably extend beyond the 3' ends of the rRNAs, creating a slight 5' overhang. The RNA population can then be treated with a polyA polymerase and ATP. RNAs with non-hybridized 3' ends can be extended by the action of the polymerase, creating a 3' polyA tail on the mRNA portion of the sample. The resulting polyA modified RNA can then be reverse transcribed using a tag comprising oligo-dT and amplification and fingerprint domains.

4. Tagging DNA

DNA (*e.g.*, genomic DNA and cDNA) can be tagged by various methods, including primer extension or ligation. In certain aspects, the DNA may be single stranded or double stranded.

a. Single Stranded DNA

In one embodiment, a target single-stranded DNA (*e.g.*, cDNA) population may be diluted in a buffer appropriate for hybridization and polymerization, and hybridized to one or more tags comprising target specific domain(s). Addition of a DNA polymerase such as, for example, the klenow fragment of DNA polymerase I or Taq DNA polymerase, will extend a tag to create a tagged population of DNA segments.

In aspects where the DNA is double stranded (*e.g.*, genomic DNA), it may be denatured prior to tagging by any of a variety of methods known in the art, including, for example, heating to 95°C in a solution of 0.2 M NaOH. In certain aspects, the denatured DNA may be removed or purified from the denaturing reagents by methods well known to those of skill in the art, such as, for example, ethanol precipitation. The denatured DNA may then be tagged using primer extension as described above or any other techniques that would be known to one of ordinary skill in the art.

b. Double Stranded DNA

In certain embodiments, double stranded DNA may be tagged by ligation. For example, a double-stranded DNA can be digested with a restriction enzyme, and one or more tags comprising a compatible restriction fragment cut site may be ligated to the digested DNA. It is particularly preferred that the unligated tags be removed prior to amplification to keep them from participating in the amplification reaction.

A disadvantage of appending double-stranded tags to double-stranded nucleic acids (*e.g.*, DNA) is that primers specific to the amplification domain of the tag can bind and be extended from target and non-target molecules alike. This is in contrast to other tagging methods described herein, whereby single-stranded tags are appended to single-stranded nucleic acids from the sample. In the embodiments employing single-stranded tags and nucleic samples, the amplification domain of the tag sequence only becomes a primer binding site when the arbitrary primer is extended

during the amplification phase. The use of single-stranded tagging is greatly preferred as it reduces background signal.

E. AMPLIFICATION

After a nucleic acid sample is tagged, it may be used to generate an amplified population comprising a set of distinct amplified nucleic acids. In preferred embodiments, the use of a fingerprint domain internal to an amplification domain is used to generate labeled nucleic acids from the amplification products to generate a nucleic acid fingerprint of the sample. Separating the amplification and fingerprint synthesis reactions can effectively eliminate most or all undesired amplification products that result from primer interactions with internal sites in a nucleic acid population. The use of an internal fingerprint domain may also eliminate undesired amplification products that result from interactions between the primers (*e.g.*, primer dimers).

It is preferred that any unincorporated and undesired tag primer be removed prior to amplification to keep the tag primers from competing with the amplification primers for templates during amplification. A primer can be removed from the sample using, for example, size exclusion chromatography (Sambrook 1989). Supports with a pore size large enough to allow the tags or adaptors to enter while excluding the larger nucleic acids provide can provide an easy way to generate primer-free nucleic acids. In some embodiments, the free tags or adaptors can be removed from a nucleic acid population by differential precipitation. For example, LiCl and ethanol are both known to preferentially precipitate larger DNA, therefore, as would be known to one of ordinary skill in the art, appropriate conditions may be developed to separate DNA from the oligonucleotide tags or adaptors prior to amplification.

Although slight variations abound, the general principle of amplifying a subset of nucleic acids from a complex sample population using arbitrary primers is the same. An RNA or DNA sample is contacted with one or more amplification primers that are able to hybridize to targets within the nucleic acid sample. Once hybridized to complementary domains in the sample nucleic acids, the primer(s) are extended by a nucleic acid polymerase (*e.g.*, a DNA polymerase) to create copies of specific sequences in the sample.

In embodiments of the present invention, a subset of nucleic acids in a tagged cDNA population may be amplified using a primer specific to the appended tag and one or more arbitrary primers. In embodiments wherein a tagged nucleic acid in a sample possesses a sequence complementary to one of the arbitrary primers in the reaction, then the fragment is exponentially amplified by the combined action of the tag and arbitrary primers. These methods may reduce very complex nucleic acid populations to between, for example, about 10 to about 1000 amplified nucleic acids.

Similar methods for generating amplified populations involve reducing the complexity of DNA populations by digesting sample DNA with one or more restriction enzymes and ligating an adapter comprising a PCR primer binding site to the digested DNA. In general embodiments, an adapter comprises a double-stranded oligonucleotide with a restriction site and a primer binding site. In certain aspects, digesting a dsDNA and the adapter with an appropriate restriction enzyme, then conducting a ligation reaction appends the primer binding site to the nucleic acid population. An adapter ligated to the digested DNA creates a limited population of DNA sequences possessing the adapter at its terminus. If the DNA being digested were tagged, the adapter-modified DNA targets could be amplified using primers specific to the adapter and the tag.

A number of template dependent processes are available to amplify sequences present in a given sample. A non-limiting example is the polymerase chain reaction (referred to as PCR™) which is described in detail in U.S. Patents 4,683,195, 4,683,202 and 4,800,159, and in Innis *et al.*, 1988, each of which is incorporated herein by reference in their entirety. Other non-limiting methods for amplification of target nucleic acid sequences that may be used in the practice of the present invention are disclosed in U.S. Patents 5,843,650, 5,846,709, 5,846,783, 5,849,546, 5,849,497, 5,849,547, 5,858,652, 5,866,366, 5,916,776, 5,922,574, 5,928,905, 5,928,906, 5,932,451, 5,935,825, 5,939,291 and 5,942,391, GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety.

In another embodiment, a reverse transcriptase PCR™ amplification procedure may be performed to amplify mRNA populations. Methods of reverse transcribing RNA into cDNA are well known (see Sambrook, 1989). Alternative methods for reverse transcription utilize thermostable DNA polymerases. These

methods are described in WO 90/07641. Additionally, representative methods of RT-PCR are described in U.S. Patent 5,882,864.

F. FINGERPRINT REACTIONS

A fingerprint reaction may comprise any of a number of methods that generate labeled nucleic acids from amplification products. In general embodiments, the fingerprint populations of different samples are compared to identify sequences that are unique to a sample or present at higher or lower levels in a sample. In preferred embodiments of the invention, a fingerprint reaction is accomplished using the fingerprint domain of appended tags.

i. Primer Extension

In certain embodiments, a fingerprint domain comprises a primer binding site internal to the amplification domain. Further, the fingerprint domain may comprise a binding site for a fingerprint primer. A fingerprint primer may be hybridized to amplified nucleic acids and extended by a DNA polymerase (*e.g.*, klenow fragment of DNA polymerase I or Taq DNA polymerase) to produce a labeled nucleic acid.

As used herein, a "labeled product" or a "labeled nucleic acid" is a nucleic acid that comprises a detectable molecule or moiety. Detectable moieties include non-isotopic reagents, isotopic reagents or combinations thereof. Non-isotopic compounds used for labeling are typically an affinity ligand such as, for example, a biotin, a digoxigenin, or a DNP, or a fluorescent dye such as Cy3 or Cy5 that are attached covalently to a primer, one or more dNTPs being incorporated during primer extension or both. Alternatively, one or more radiolabeled atoms (*e.g.*, ^{32}P, ^{33}P, or ^{35}S) may be incorporated into the primer, nucleotides or both. Of course, other detectable moieties that would be known to those of skill in the art in light of the disclosures herein may be used.

In some aspects, a fingerprint primer can be hybridized to an amplified population generated from a single sample as a sample mixture. In embodiments wherein labeled nucleotides are being incorporated using the primer extension method, it is preferred to keep an amplification primer from being extended and creating labeled DNAs from all of the amplified nucleic acid(s), as this would increase false positives.

Thus, in particularly preferred aspects, the non-fingerprint primers are removed from amplified nucleic acids prior to initiating a fingerprint reaction. A

primer can be removed using techniques that would be known to those of skill in the art, such as for example, size exclusion chromatography or precipitation of nucleic acids using conditions that keep primers in solution (Sambrook, 1989). For example, a nucleic acid population can be added to a size exclusion column and centrifuged. The population collects in the filtrate, free of the column-bound amplification primers.

In cases where an amplified sample mixture is used to generate labeled fingerprints specific to the samples comprising the sample mixture and labeled dNTPs are being incorporated by primer extension, the amplified nucleic acids may be split into two or more different labeling reactions, up to the number of reactions as there were sample populations. In certain aspects, the labeling reactions are made specific to each fingerprint domain of the sample by using primers that only anneal to one of the unique fingerprint domains. Labeled DNA resulting from each fingerprint reaction are specific to each sample.

In cases where labeled primers are used to generate labeled DNA from the amplification products from a sample mixture, a single reaction can be used to differentially label fingerprints specific to each of the various samples. For example, if the fingerprint primer used to label amplification products from one sample has Cy3 and the fingerprint primer for the second sample has Cy5, then the two primers could be added to the amplification products generated from a sample mixture. The fingerprint primers would hybridize to the appropriate amplification products and be extended by the action of a DNA polymerase. Targets derived from one sample would be labeled exclusively with Cy3 while targets from the second sample would be labeled with Cy5. The labeled nucleic acids could then be fractionated and detected in a way that the signals from Cy3 and Cy5 could be distinguished.

ii. Differential Labeling by *In vitro* Transcription

In embodiments wherein the fingerprint domain comprises a promoter, a transcription reaction with one or more labeled nucleotides (*e.g.*, isotopic- or non-isotopic-labeled NTPs) and an appropriate RNA polymerase can be used to convert double-stranded templates into nucleic acid fingerprints that can be used for comparative analysis. In certain embodiments, a DNA population can be converted to labeled RNAs by *in vitro* transcription.

RNA polymerases are well known to those of ordinary skill in the art. For example, several phage RNA polymerases have been isolated and characterized (Sambrook 1989). Additional RNA polymerases may be isolated from nature or by a mutation/selection screen using an existing polymerase (Ikeda 1993). Any such polymerases or promoters are contemplated for use in the present invention.

2. Affinity sites

A fingerprint domain of a tag may comprise a sequence with an affinity for a specific nucleic acid, protein, or other binding ligand. A binding ligand may comprise, but is not limited to, an oligonucleotide complementary to a fingerprint domain, a nucleic acid binding protein (*e.g.*, a transcription factor) that binds a specific DNA or RNA sequence, a small molecule that intercalates into a given RNA or DNA sequence or combinations thereof. A binding ligand may either be bound to a solid support (*e.g.*, a single bead or a membrane) or otherwise readily removed or separated from a solution.

In certain embodiments, the methods of the present invention may comprise a labeling step to provide labeled nucleic acids in or from the amplification products generated from tag and arbitrary or adapter priming. In specific aspects, a population would then be applied to a solution or solid support possessing a ligand specific to the fingerprint domain, the tagged sample or one of the samples if a sample mixture was amplified. The specifically isolated targets could then be used for fingerprint analysis.

G. FINGERPRINT ANALYSIS

Once labeled nucleic acids that comprise a fingerprint have been generated, the individual nucleic acids may be fractionated to foster analysis. In some embodiments, nucleic acids are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis, capillary electrophoresis, column chromatography, or any other of the methods commonly employed that differentiate nucleic acids based on size differences. In particular aspects, labeled nucleic acid fingerprints from different samples can be compared in adjacent lanes using any technique described herein (*e.g.*, by gel electrophoresis, autoradiograpy) and/or such methods that would be known to one of ordinary skill in the art.

In other embodiments, separation of nucleic acids may also be conducted by chromatographic techniques known in the art. There are many kinds of

chromatography which may be used in the practice of the present invention, including adsorption, partition, ion-exchange, hydroxylapatite, molecular sieve, reverse-phase, column, paper, thin-layer, and gas chromatography as well as HPLC. The number, location and relative abundance of the bands resulting from such amplification and/or fingerprint reactions is the "fingerprint" of the nucleic acid sample.

In some embodiments, separation of nucleic acids may be conducted by hybridization methods (Trenkle et al, 1999). For example, the fingerprint population may be hybridized to an array which will fractionate the fingerprint nucleic acids according to their sequences. The intensity of signal on the spots of the array provides the fingerprint for the sample population.

In certain aspects, the detection of a fingerprint population may be performed by visual means (Sambrook *et al.,* 1989). Alternatively, the detection may involve indirect identification of the population by chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent label or even by a system using electrical and/or thermal impulse signals (Affymax technology; Bellus, 1994). In certain embodiments, the nucleic acids are visualized. For instance, if the nucleic acids are integrally labeled with radio- or fluorometrically-labeled nucleotides, the separated labeled nucleic acids can be exposed to x-ray film or visualized under the appropriate excitatory spectra.

Other methods of nucleic acid detection that may be used in the practice of the instant invention are disclosed in U.S. Patent No. 5,840,873, which describes detection of multiple nucleic acids utilizing oligonucleotide probes coupled to different chemiluminescent labeling reagents; U.S. Patent 5,843,651, which describes an method for detecting nucleic acid sequence differences through binding to a ligand; U.S. Patent 5,846,708 describes an array based binding and detection system; U.S. Patent 5,846,717, describes detection systems based on nucleic acid cleavage; U.S. Patent 5,846,726, describes nucleic acid probes that detects nucleic acids by changes in fluorescence of the a probe attached dye after cleavage of a nucleic acid; U.S. Patent 5,846,729 describes nucleic acid detection in solution using changes in fluorescence of probes; U.S. Patent 5,849,487 describes use of a nuclease resistant probe to detect a nucleic acid sequence; and U.S. Patent 5,905,024 multi-step fractionation of nucleic acid by binding to arrays, each of which is incorporated herein by reference.

In certain embodiments, individual nucleic acids comprising a fingerprint band may be cut out and eluted from the gel for further manipulation.

Several methods have been described herein for reducing the complexity of DNA and RNA populations for performing comparative analyses. Those that would benefit from the tagging methods of the present invention include DD-RT-PCR, AP-PCR, and restriction digest mediated amplification fingerprinting. The invention can be applied to improve either of these general technologies, as described in the Examples below.

H. IDENTIFYING A TAG

Where co-amplification is being used to assess nucleic acids, unique tags are used for different sample populations. It is important that the unique tags not contribute to amplification or fingerprint biases (e.g., differences in amplification or fingerprint efficiencies). New tag sequences should be tested to ensure that they function equivalently. One such test involves splitting a single sample into separate tagging reactions incorporating the different tags. After tagging, the samples are mixed, amplified, and used to synthesize fingerprints unique to each tagged sample. The nucleic acid fingerprints are assessed using the method of analysis that is to be employed. If the number of bands, the migration of the bands, or the intensity of the bands varies in the analysis of the nucleic acid fingerprints, then the tags are not functioning equivalently.

Identifying fingerprint domains that function equally well and that do not affect amplification efficiency is relatively straightforward where primer extension, affinity purification or digestion is being used to generate fingerprints. In these cases, altering the identity of just a few nucleotides can provide effective synthesis of fingerprints specific to each sample. Rarely does altering a few bases within the fingerprint domain affect amplification efficiency. In addition, because both methods use the same enzyme (*i.e.*, a single DNA polymerase) for generating labeled nucleic acids from each of the unique tags, polymerization biases should not introduce variability.

However, where *in vitro* transcription is used to synthesize nucleic acid fingerprints, bias is far more likely to occur. Promoters for the well-characterized phage RNA polymerases are similar in base content, but they stretch over 15-20 nucleotides creating a relatively large, unique sequence domain within the amplified

nucleic acids. In addition, different RNA polymerases are likely to possess sequence biases that affect transcription efficiency that might affect synthesis of the fingerprint. This has not affected the method of the present invention in the examples conducted and described herein. However, it is possible that this may affect certain embodiments. To overcome these potential problems, mutants of a single RNA polymerase that do not affect enzymatic activity but do alter promoter specificity may be used in the methods of the present invention may be designed (Ikeda 1993). This methodology may allow the creation of promoter sequences and/or mutant polymerases that provide equal amplification efficiency to be used to distinguish differentially tagged amplified nucleic acids.

J. EXAMPLES

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1

REDUCING AMPLIFICATION ARTIFACTS USING POPULATION TAGS

FIG. 1 illustrates the benefits of the invention. Amplification of a nucleic acid sample with arbitrary and tag specific primers results in four basic types of amplification product. Those amplification products that derive from the tagged ends of sample nucleic acids and those that derive from other regions of the nucleic acids in the sample due to the action of two arbitrary primers, two tag primers, or one tag primer and one arbitrary primer. The reproducibility of the latter three amplification products is relatively low, making them a primary source of false positives in methods incorporating fingerprint analysis. The present invention incorporates a labeling step subsequent to amplification that requires the fingerprint domain. Thus amplification products that do not derive from the terminus of the nucleic acid sample are not visible in the fingerprint analysis.

An example of a general makeup of a tag relative to the population being assayed is provided in FIG. 2. The thin line represents the RNA or DNA molecules from a nucleic acid sample. The bold line represents the appended tag. F.D. is the fingerprint domain of the tag. A.D. is the amplification domain of the tag. Arb. Seq. is a position within the sample population that is bound and extended by an arbitrary primer. As indicated in FIG. 2, it is particularly preferred that the fingerprint domain is between an amplification domain and the sequence of each nucleic acid target of the population to be amplified. In other words, it is particularly preferred that a fingerprint domain is internal to the amplification domain. Depending on the method used for labeling, the labeling and amplification domains may or may not be present in the final fingerprint population.

EXAMPLE 2

TAG SEQUENCES APPENDED VIA REVERSE TRANSCRIPTION

This example describes one embodiment wherein a tag comprising a fingerprint domain that includes a promoter (*e.g.*, either a T7 or SP6 promoter) was appended to cDNA using reverse transcription.

1 μg of thymus total RNA or liver total RNA was mixed with 0.5 mM dNTPs, 50 μM T7-VN-1 or SP6-VN-1, and reaction buffer (50 mM Tris pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT). The T7-VN-1 and SP6-VN-1 reverse transcription primers were are as follows:

T7-VN-1

GCTGATGGCGATGAATGAACACTG*TAATACGACTCACTATAGGGAG*ATTT TTTTTTTTTVN (SEQ ID NO:1).

SP6-VN-1

GCTGATGGCGATGAATGAACACTG*ATTTAGGTGACACTATAGAAGAG*TTTT TTTTTTTTVN (SEQ ID NO:2).

In regard to these and all of the other reverse transcription primer sequences disclosed in this specification:

Bold = Amplification Domain,

Italics = Fingerprint Domain, and

Plain Text = Anchored oligodT.

The mixture was heated to 68°C for five minutes, then cooled to 42°C. RNase inhibitor (40 units/μl) and Moloney-Murine Leukemia Virus-Reverse Transcriptase (M-MLV-RT) (20 units/μl) were added and the mixture was incubated at 42°C for one hour to convert the RNA samples to tagged cDNAs.

To remove the reverse transcription primers, the samples were applied to S-200 HR sephacryl spin columns. The columns were spun at 700 × G for 2 minutes. The filtrate was recovered, providing the tagged cDNA population used for subsequent analyses.

EXAMPLE 3

TAG SEQUENCES APPENDED VIA CAP –DEPENDENT LIGATION

This example describes how an RNA can be tagged via cap-dependent ligation. Of course, those of skill in the art will understand that there are additional manners of appending tag sequences, as disclosed in this specification and/or known in the art.

1 μg of mouse thymus or liver total RNA was treated with 1 unit of Calf Intestine Phosphatase in a 20 μl reaction of 50 mM Tris pH 8.5 and 0.1 mM EDTA to remove 5'terminal phosphates from uncapped RNAs. The reaction was incubated at 37°C for 1 hr. The volume of the reaction was increased to 150 μl by the addition of 500 mM Ammonium acetate, 1 mM EDTA. One acid phenol/chloroform and one chloroform extraction were done. The RNA was then precipitated and air dried.

To convert the capped mRNA in the sample to RNA with 5'monophosphates, the dephosphorylated RNA population was dissolved in 8 μl of water. 1 μl of 10X TAP buffer (500 mM NaOAc, 10 mM EDTA, 1% beta-mercaptoethanol, 0.1% CHAPS) and 0.1 unit of TAP were added and the decapping reaction was allowed to proceed for 1 hour at 37°C. Ammonium acetate was added to the reaction to provide a final concentration of 0.5 M. A two-fold volume of ethanol was added and the RNA was precipitated by incubation at -20°C for fifteen minutes. The precipitated RNA was recovered by centrifugation.

To ligate the primers to the RNA possessing a 5' monophosphate, the RNA pellet was dissolved in 8 μl of water. 1 μl of 10X ligation buffer (500 mM Tris pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 1 mM ATP), 1 μg of the ligation tag, and 5 units of T4 RNA ligase were added. The ligation tag sequence is shown, with the

amplification domain in bold and the fingerprint domain in italics: 5'TAATACGACTCACTATAGGGTTCGGGCTTAGGCTCCAGTGCCTGTTCGGT GGTCGCGGC**GCTGATGGCGATGAATGAACACTG**CGGCAAGCCGCTTAA T*GACACTCGTTTGCTGGCTTTGATG*GGCGAGCTGGAAGGCCGTATCTCCGGC AGCATTCATTACGACAAA 3' (SEQ ID NO:3). The ligation reaction was allowed to proceed for 1 hour at 37°C.

To remove the unincorporated ligation tags, the samples were applied to S-200 HR sephacryl spin columns. The columns were spun at 700 × G for 2 minutes. The filtrate was recovered, providing the tagged cDNA population used for subsequent analysis.

EXAMPLE 4

GENERATING LABELED FINGERPRINTS FROM TAGGED AMPLIFICATION PRODUCTIONS BY PRIMER EXTENSION

This example describes reverse transcription of tagged RNA, amplification and differentiation via primer extension. In this embodiment, labeling by primer extension according to the invention employs a domain within the tag sequence that possess sufficient sequence to direct specific primer hybridization.

After removal of the ligation tags, the ligated RNA of Example 3 was mixed with 0.5 mM dNTPs, 50 μM random sequence decamers, reaction buffer (50 mM Tris pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT). The mixture was heated to 68°C for five minutes, then cooled to 42°C. RNase inhibitor (40 units/μl) and M-MLV-RT (20 units/μl) were added and the mixture was incubated at 42°C for one hour to convert the tagged RNA sample to tagged cDNA.

To remove the RT primers, the samples were applied to S-200 HR sephacryl spin columns. The columns were spun at 700 × G for 2 minutes. The filtrate was recovered, providing the tagged cDNA population used for subsequent analysis.

A subset of the tagged cDNA population was amplified by PCR™. The PCR™ comprised 1 μl of the cDNA filtrate, with 0.2 μM Tag primer, [GCTGATGGCGATGAATGAACACTG (SEQ ID NO:4)], 0.2 μM Arbitrary primer 1: [AGCAGTTGGTTGGAGCAACATC (SEQ ID NO:5)], 500 μM dNTPs, reaction buffer (75 mM KCl, 50 mM TRIS pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. The first cycles of PCR™ used the following profile: 94 °C, 5 minutes;

then 5 cycles of: 94 °C, 30 seconds; 42 °C, 30 seconds; 72 °C, 60 seconds. Twenty-five additional cycles were performed using 58°C as the annealing temperature rather than 42°C. At the end of the final amplification cycle, a 5 minute, 72°C soak was used for annealing and primer extension.

To remove the amplification primers, the samples were applied to S-200 HR sephacryl spin columns. The columns were spun at 700 × G for 2 minutes. The filtrate was recovered, providing the tagged cDNA population used for subsequent analysis.

A labeling reaction was conducted using the fingerprint domain of the tag sequence. The amplified nucleic acid was added to a linear amplification reaction comprising 100 μM dGTP, 100 μM dTTP, 100 μM dCTP, 10 μM dATP, 0.5 μM [α-^{32}P]-dATP, reaction buffer (75 mM KCl, 50 mM TRIS pH 9.0, 2 mM $MgCl_2$), 2.5 units SuperTaq polymerase, and 0.2 μM of the following fingerprint primer: GACACTCGTTTGCTGGCTTTGATG (SEQ ID NO:6).

Five cycles of the following PCR™ program were used to extend the primer multiple times on the amplified nucleic acid: 94 °C, 5 minutes; then 5 cycles of: 94 °C, 30 seconds; 55 °C, 30 seconds; 72 °C 60 seconds. The radiolabeled nucleic acids were fractionated on a denaturing polyacrylamide sequencing gel to create a fingerprint representative of the input RNA sample.

Autoradiography revealed a fingerprint of approximately 25 bands of varying intensity.

EXAMPLE 5

GENERATING LABELED FINGERPRINTS FROM TAGGED POPULATIONS BY TRANSCRIPTION

This example demonstrates that when the fingerprint domains of tags include a transcription promoter, amplified populations can be used to generate a labeled RNA fingerprint for comparative expression analysis. The protocols described in Example 2 were used to supply cDNA for amplification.

A subset of the cDNA population was amplified by mixing 1 μl of the cDNA filtrates produced in Example 1 with the following primers: Tag primer: GCTGATGGCGATGAATGAACACTG (SEQ ID NO:7); Arbitrary primer 1:

AGCAGTTGGTTGGAGCAACATC; Arbitrary primer 2: CAGGGACCTTCTCTAAGACTTTC (SEQ ID NO:8).

The Tag primer was coupled with Arbitrary primer 1 in the first reaction and with Arbitrary primer 2 in the second reaction. The PCR™ comprised 0.2 μM of the indicated primers, 500 μM dNTPs, reaction buffer (75 mM KCl, 50 mM Tris pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. The first cycles of PCR™ used the following profile: 94 °C, 5 minutes; then 5 cycles of: 94 °C, 30 seconds; 42 °C, 30 seconds; 72 °C, 60 seconds. Twenty-five additional cycles were performed using 58°C as the annealing step rather than 42°C. At the end of the final amplification cycle, a 5 minute, 72°C soak was used for annealing and primer extension.

Transcription reactions were then used to convert the amplified nucleic acids into RNAs for fingerprint analysis. 2 μl aliquots of the PCR™ samples were used to generate labeled RNAs. Transcription reactions comprising 10 U/μl of T7 or SP6 RNA polymerase, reaction buffer (20 mM NaCl, 40 mM TRIS pH 8, 6 mM $MgCl_2$, 2 mM spermidine HCl, and 10 mM DTT), 0.5 mM ATP, GTP, and CTP, 10 μM UTP, and 2.5 μM [α-^{32}P]-UTP. The reactions were incubated at 37°C for one hour. The use of the two polymerases depended on the identity of the tag used for reverse transcription. The cDNAs primed by T7-VN-1 were transcribed with T7 RNA polymerase. The cDNAs primed by SP6-VN-1 were transcribed with SP6.

Following transcription, the labeled RNAs were prepared for electrophoresis by addition of a 2X volume of gel loading buffer (95% Formamide/0.025% Xylene Cyanol/0.025% Bromophenol blue/0.5 mM EDTA/0.025% SDS) and heating to 95°C for five minutes. The labeled RNAs were then fractionated by gel electrophoresis using a 6% acrylamide/7 M urea gel.

Autoradiography revealed a fingerprint of approximately 25 bands of varying intensity. Approximately 10% of the bands varied between the thymus and liver samples.

EXAMPLE 6

TAGGED DD-RT-PCR

Standard DD-RT-PCR involves reverse transcribing an mRNA population with anchored oligo-dT. FIG. 3 demonstrates an embodiment of the present invention that uses a tagged oligodT primer and an arbitrary sequence primer to generate a

fingerprint in a manner similar to DD-RT-PCR. In such a protocol, two or more populations can be compared by performing reverse transcription with anchored oligodT primers possessing identical PCR™ primer binding sites and different transcription promoters (*e.g.*, T7 RNA polymerase for one, SP6 in the other).

FIG. 3, describes this protocol generally. Anchored oligodT primers, each including a 5' end amplification domain (*e.g.*, a PCR™ primer binding site (PBS)), a fingerprint domain (for example, a promoter for T7 or SP6 RNA polymerase), and anchored oligodT (i.e., TnVN) prime reverse transcription. Reverse transcription of at least two populations of RNA with the primers generates respective cDNA population(s) with both a PCR™ primer binding site and a transcription promoter at the 5' end of every cDNA. Subsets of the resulting populations can then be amplified by PCR™ with a primer specific to the amplification domain and an arbitrary primer. The amplified populations are then used to initiate transcription reactions that include reagent(s) that produce labeled RNA from the fingerprint domain (*e.g.*, the T7 or SP6 RNA polymerase and a labeled NTP such as [α-^{32}P]-UTP). The resulting labeled RNA is fractionated (*e.g.*, by gel electrophoresis or array hybridization) and visualized to create a fingerprint of the RNA in each population. This fingerprints of the populations are then compared. The benefit is that only amplified nucleic acids that include the anchored 5' end of the cDNA are displayed, reducing the impact of unpredictable primer interactions. A specific study employing the general methods described above was conducted as follows.

1 μg of thymus total RNA or liver total RNA was mixed with 0.5 mM dNTPs, 50 μM T7-VN-1, and reaction buffer (50 mM Tris pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT). The mixture was heated to 68°C for five minutes, then cooled to 42°C. RNase inhibitor (40 units/μl) and M-MLV-RT (20 units/μl) were added and the mixture was incubated at 42°C for one hour to generate tagged cDNA samples.

To remove the unincorporated tag primers, the samples were applied to S-200 HR sephacryl spin columns. The columns were spun at 700 × G for 2 minutes. The filtrate was recovered, providing the tagged cDNA population used for subsequent analysis.

Amplification was conducted by adding 1 μl of the cDNA filtrates to a combination of the following amplification primers: Tag primer:

GCTGATGGCGATGAATGAACACTG, Arbitrary primer 1: AGCAGTTGGTTGGAGCAACATC; Arbitrary primer 2: CAGGGACCTTCTCTAAGACTTTC. The Tag primer was coupled with Arbitrary primer 1 in the first reaction and with Arbitrary primer 2 in the second reaction. The PCR™ comprised 0.2 μM of the indicated primers, 0.5 mM dNTPs, reaction buffer (75 mM KCl, 50 mM TRIS pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. The first cycles of PCR™ used the following profile: 94 °C, 5 minutes, then 5 cycles of: 94 °C, 30 seconds, 42 °C, 30 seconds, 72 °C, 60 seconds. Twenty-five additional cycles were performed using 58°C as the annealing step rather than 42°C. At the end of the final amplification cycle, a 5 minute, 72°C soak was used for annealing and primer extension.

2 μl aliquots of the PCR™ samples were used to generate labeled RNA for fingerprint analysis. Transcription reactions comprising 10 U/μl of T7 RNA polymerase, reaction buffer (20 mM NaCl, 40 mM Tris pH 8, 6 mM $MgCl_2$, 2 mM spermidine HCl, and 10 mM DTT), 0.5 mM ATP, GTP, and CTP, 10 μM UTP, and 2.5 μM [α-^{32}P]-UTP. The reactions were incubated at 37°C for one hour.

Following transcription, the labeled RNAs were prepared for electrophoresis by addition of a 2X volume of gel loading buffer (95% Formamide/0.025% Xylene Cyanol/0.025% Bromophenol blue/0.5 mM EDTA/0.025% SDS) and heating to 95°C for five minutes. The labeled RNAs were then fractionated by gel electrophoresis using a 6% acrylamide/7 M urea gel providing fingerprints for the liver and thymus RNAs.

Autoradiography revealed approximately 25 district bands from both liver and thymus RNA using either arbitrary primer. Approximately 10% of the products varied between thymus and liver from both arbitrary primers.

EXAMPLE 7

CO-AMPLIFIED, TAGGED DD-RT-PCR

This example demonstrates an embodiment of the present invention that uses tagged cDNA samples in a variation of the DD-RT-PCR procedure whereby two samples are co-amplified to enhance comparative analysis.

FIG. 4 depicts this embodiment. Reverse transcription primers, each including an amplification domain, a fingerprint domain, and anchored oligodT, as described for

FIG. 3, are used in the reverse transcription of at least two populations of RNA. The fingerprint domains are different for each of the nucleic acid samples being tagged. For instance, one fingerprint domain can be a promoter for T7 RNA polymerase and one can be a promoter for SP6 RNA polymerase. The reverse transcription generates respective cDNA population(s) with both a PCR™ primer binding site and a transcription promoter at the 5' end of every cDNA. The tagged cDNA samples are then mixed and co-amplified with an arbitrary primer and a primer specific to the amplification domain of the tags. The resulting amplified DNA is then split into multiple transcription reactions (*e.g.*, one with T7 RNA polymerase and one with SP6 RNA polymerase), which include reagent(s) that produce labeled RNA from amplification products bearing an appropriate fingerprint domain. The resulting labeled RNA is fractionated (*e.g.*, by gel electrophoresis or array hybridization) and visualized to create a fingerprint of the RNA in each population. Fingerprints of the populations are then compared. In addition to subtracting non-anchor-derived amplification products from the fingerprint population, the co-amplification procedure obviates the impact of tube-to-tube variation that is common in nucleic acid amplification reactions.

In a specific study demonstrating the techniques described in this Example, RNA samples were each tagged with different tags. One tag possessed a fingerprint domain that is a promoter for T7 RNA polymerase and the other had a promoter for SP6 RNA polymerase. 1 μg of thymus total RNA or liver total RNA was mixed with 0.5 mM dNTPs, 50 μM T7-VN-1 or SP6-VN-1, and reaction buffer (50 mM Tris pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT). The mixtures were heated to 68°C for five minutes, then cooled to 42°C. RNase inhibitor (40 units/μl) and M-MLV-RT (20 units/μl) were added and the mixtures were incubated at 42°C for one hour. The four reactions generated the following tagged cDNA populations: thymus cDNA with a T7 promoter, thymus cDNA with an SP6 promoter, liver cDNA with a T7 promoter, and liver cDNA with an SP6 promoter.

To remove the unincorporated tag primers, the four cDNA populations were applied to separate S-200 HR sephacryl spin columns. The columns were spun at 700 × G for 2 minutes. The filtrates were recovered, providing the tagged cDNA populations used for subsequent amplification. Four mixtures of the cDNA populations were generated for subsequent analysis. The tagged cDNAs were

designated as thymus T7, thymus SP6, liver T7, and liver SP6. The mixtures were equal amounts of thymus T7 and thymus SP6, thymus T7 and liver SP6, thymus SP6 and liver T7, and liver T7 and liver SP6.

1 μl of each of the cDNA mixtures was used to initiate amplification using combinations of the following primers: Amplification primers: Tag primer: GCTGATGGCGATGAATGAACACTG; Arbitrary primer 1: AGCAGTTGGTTGGAGCAACATC; Arbitrary primer 2: CAGGGACCTTCTCTAAGACTTTC.

The Tag primer was coupled with Arbitrary primer 1 in four reactions for the four different sample mixtures. Four additional reactions with the four different sample mixtures coupled the Tag primer with Arbitrary primer 2. The PCR™s comprised 0.2 μM of the indicated primers, 500 μM dNTPs, reaction buffer (75 mM KCl, 50 mM Tris pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. The first cycles of PCR™ used the following profile: 94 °C, 5 minutes; then 5 cycles of: 94 °C, 30 seconds; 42 °C, 30 seconds; 72 °C, 60 seconds. Twenty-five additional cycles were performed using 58°C as the annealing step rather than 42°C. At the end of the final amplification cycle, a 5 minute, 72°C soak was used for annealing and primer extension.

2 μl aliquots of each of the PCR™ samples were added to transcription reactions with T7 RNA polymerase or SP6 RNA polymerase. The transcription reactions contained 10 U/μl of T7 or SP6 RNA polymerase, reaction buffer (20 mM NaCl, 40 mM TRIS pH 8, 6 mM $MgCl_2$, 2 mM spermidine HCl, and 10 mM DTT), 0.5 mM ATP, GTP, and CTP, 10 μM UTP, and 2.5 μM [α-32P]-UTP. The reactions were incubated at 37°C for one hour. The sixteen labeled RNAs all derived from the two input RNA samples and the four cDNA samples (two tags per RNA).

Fingerprint analysis was conducted following transcription. The labeled RNAs were prepared for electrophoresis by addition of a 2X volume of gel loading buffer (95% Formamide/0.025% Xylene Cyanol/0.025% Bromophenol blue/0.5 mM EDTA/0.025% SDS) and heating to 95°C for five minutes. The labeled RNAs were then fractionated by gel electrophoresis using a 6% acrylamide/7 M urea gel providing fingerprints. The two samples derived from each amplification reaction were assessed in adjacent lanes.

As expected, where the two liver cDNAs and the two thymus cDNAs were co-amplified, the labeled fingerprints were identical. Also as expected, the fingerprints derived from the amplification reactions used to co-amplify the liver and thymus RNAs varied at several locations. These data demonstrate that the present invention reduces the rates of false positives, making nucleic acid fingerprint analysis more robust.

EXAMPLE 8

TAGGED AP-PCR

This example demonstrates that the present invention can be applied to the Arbitrarily-Primed Polymerase Chain Reaction (AP-PCR) procedure.

AP-PCR requires that an mRNA population be reverse transcribed and amplified using arbitrary sequence primers. If the primer used for reverse transcription includes at its 5' end an amplification domain (*e.g.*, a PCR™ primer binding site) and a fingerprint domain (*e.g.*, a promoter for T7 RNA polymerase), the reverse transcription will generate a cDNA population with both a PCR™ primer binding site and a transcription promoter at the 5' end of every cDNA. A subset of the resulting population can be amplified by PCR™ with a primer specific to the amplification domain and an arbitrary primer. The amplified population can then be used to initiate a transcription reaction that includes reagent(s) that produce labeled nucleic acids from the fingerprint domain. The resulting labeled nucleic acids can be fractionated (*e.g.*, by gel electrophoresis or array hybridization) and visualized to create a fingerprint for the sample. This fingerprint can be compared to the fingerprints of other samples generated using the same procedure. The benefit is that only amplified nucleic acids that include the 5' end of the cDNA are displayed, reducing the impact of unpredictable primer interactions.

A collection of RNAs possessing sequence derived from hIL-1B, hIL-2, hIL-3, hIL-4, hIL-8, and hIL-10 were used to assess the feasibility of using tagged cDNAs with the AP-PCR procedure. The RNAs were encapsidated by the MS2 coat protein. Each of the RNAs shared sequence at their 5' and 3' ends, making it possible to reverse transcribe and amplify all of the targets with a single set of primers. The amplified nucleic acids may be distinguished based on the different gene specific sequences.

The reverse transcription primer was T7-OP-1:

GCTGATGGCGATGAATGAACACTG*TAATACGACTCACTATAGGGAGA*ACA TGGGTAATCCTCATGTTT (SEQ ID NO:9).

In this primer, the underlined sequence is complementary to the target RNA in the sample, proving a priming site for tagged cDNA synthesis. The amplification domain is in bold and the fingerprint domain is italics.

The reverse transcription reactions were conducted using two RNA populations. The relative abundance of each target is indicated in the table. X is approximately equal to 10^{10} copies of the indicated RNA.

RNA	Population I	Population II
HIL-1B	1X	2X
HIL-2	1X	0X
HIL-3	1X	4X
HIL-4	1X	0.5X
HIL-8	1X	0.25X
HIL-10	1X	1X

In separate tubes, Population I and Population II were mixed with 0.5 mM dNTPs, 25 μM T7-OP-1, and reaction buffer (50 mM Tris pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT). The mixtures were heated to 68°C for five minutes, then cooled to 42°C. RNase inhibitor (40 units/μl) and M-MLV-RT (20 units/μl) were added and the mixtures were incubated at 42°C for one hour. The two reactions generated cDNAs of the input RNA possessing the promoter for T7 RNA polymerase and a PCR™ primer binding site at their 5' ends.

For removal of the unincorporated tag primers, the two cDNA populations were applied to separate S-200 HR sephacryl spin columns. The columns were spun at 700 × G for 2 minutes. The filtrate was recovered, providing the tagged cDNA populations used for subsequent amplification.

In amplification, 1 μl of each of the cDNA mixtures was used to initiate amplification using the following primers: Tag Primer: GCTGATGGCGATGAATGAACA (SEQ ID NO:10); Arbitrary primer: CCTCAGCAATCGCAGCAAACT (SEQ ID NO:11).

The PCR™s consisted of 0.2 μM of the tag and arbitrary primers, 500 μM dNTPs, reaction buffer (75 mM KCl, 50 mM Tris pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. PCR™ used the following profile: 94 °C ,3 minutes;

then 34 cycles: 94 °C, 30 seconds; 58 °C, 30 seconds; 72 °C, 30 seconds; final extension: 72 °C, 3 minutes.

For the labeling reactions, 2 μl aliquots of the two PCR™ samples were added to transcription reactions with T7 RNA polymerase. The transcription reactions contained 10 U/μl of T7 RNA polymerase, reaction buffer (20 mM NaCl, 40 mM Tris pH 8, 6 mM $MgCl_2$, 2 mM spermidine HCl, and 10 mM DTT), 0.5 mM ATP, GTP, and CTP, 10 μM UTP, and 2.5 μM [α-^{32}P]-UTP. The reactions were incubated at 37°C for one hour.

To do fingerprint analysis, the labeled RNAs were prepared for electrophoresis by addition of a 2X volume of gel loading buffer (95% Formamide/0.025% Xylene Cyanol/0.025% Bromophenol blue/0.5 mM EDTA/0.025% SDS) and heating to 95°C for five minutes. The labeled RNAs were then fractionated by gel electrophoresis using a 6% acrylamide/7 M urea gel providing fingerprints. As expected, the relative abundance of the various fingerprints varied with the amount of input RNA for each target.

EXAMPLE 9

CO-AMPLIFIED, TAGGED AP-PCR

The RNA populations were also used to demonstrate an embodiment of the present invention comprising amplified, tagged AP-PCR, incorporating co-amplification.

The reverse transcription primers were T7-OP-1, as disclosed in Example 8, and SP6-OP1:

A.D. F.D.

GCTGATGGCGATGAATGAACACTG*ATTTAGGTGACACTATAGAA*GACACA TGGGTAATCCTCATGTTT (SEQ ID NO:12).

In separate tubes, Population I and Population II were mixed with 0.5 mM dNTPs and reaction buffer (50 mM Tris pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT). 25 μM T7-OP-1 was added to the reaction with Population I and 25 μM SP6-OP-1 was added to the reaction with Population II. The mixtures were heated to 68°C for five minutes, then cooled to 42°C. RNase inhibitor (40 units/μl) and M-MLV-RT (20 units/μl) were added and the mixtures were incubated at 42°C for

one hour. The two reactions generated cDNAs of the input RNA possessing the transcription promoters and a PCR™ primer binding site at their 5' ends.

To remove the unincorporated tag primers, the two cDNA populations were applied to separate S-200 HR sephacryl spin columns. The columns were spun at 700 x G for 2 minutes. The filtrate was recovered, providing the tagged cDNA populations used for subsequent amplification.

1 µl of each of the cDNA mixtures was mixed and used to initiate amplification using the following primers: Tag Primer: GCTGATGGCGATGAATGAACA; Arbitrary primer: CCTCAGCAATCGCAGCAAACT.

The PCR™ consisted of 0.2 µM of the tag and arbitrary primers, 500 µM dNTPs, reaction buffer (75 mM KCl, 50 mM Tris pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. PCR™ used the following profile: 94 °C, 3 minutes; then 34 cycles: 94 °C, 30 seconds; 58 °C, 30 seconds; 72 °C, 30 seconds; final extension: 72 °C, 3 minutes.

2 µl aliquots of the PCR™ sample were added to transcription reactions with T7 or SP6 RNA polymerase to generate labeled RNA for fingerprint analysis. The transcription reactions contained 10 U/µl of T7 or SP6 RNA polymerase, reaction buffer (20 mM NaCl, 40 mM TRIS pH 8, 6 mM $MgCl_2$, 2 mM spermidine HCl, and 10 mM DTT), 0.5 mM ATP, GTP, and CTP, 10 µM UTP, and 2.5 µM [α-^{32}P]-UTP. The reactions were incubated at 37°C for one hour.

The labeled RNAs were prepared for electrophoresis by addition of a 2X volume of gel loading buffer (95% Formamide/0.025% Xylene Cyanol/0.025% Bromophenol blue/0.5 mM EDTA/0.025% SDS) and heating to 95°C for five minutes. The labeled RNAs were then fractionated by gel electrophoresis using a 6% acrylamide/7 M urea gel providing fingerprints. As expected, the relative abundance of the various fingerprints varied with the amount of input RNA for each target.

EXAMPLE 10

TAGGED RESTRICTION DIGEST MEDIATED AMPLIFICATION

Tagging can be applied to fingerprinting procedures that involve adapter ligation via restriction digestion. These procedures typically involve reverse transcribing an RNA sample with an anchored oligodT primer, generating ds cDNA,

digesting the sample with a restriction enzyme, ligating an adapter to the restriction site, and amplifying the population with primers specific to the adapter and oligodT. If the anchored oligodT primer includes at its 5' end an amplification domain (*e.g.*, a PCR™ primer binding site) and a fingerprint domain (*e.g.*, a promoter for T7 RNA polymerase), the reverse transcription will generate a cDNA population with both a PCR™ primer binding site and a transcription promoter at the 5' end of every cDNA. The adapter modified population can be amplified by PCR™ with a primer specific to the amplification domain and a primer specific to the ligated adapter. The amplified population can then be used to initiate a transcription reaction that includes T7 RNA polymerase and a labeled NTP (*e.g.*, [α-^{32}P]-UTP). The resulting labeled RNA can be fractionated (*e.g.*, by gel electrophoresis or array hybridization) and visualized to create a fingerprint for the sample. This fingerprint can be compared to the fingerprints of other samples generated using the same procedure. The benefit is that only amplified nucleic acids that include the anchored 5' end of the cDNA are displayed, reducing the impact of unpredictable primer interactions.

EXAMPLE 11

TAGGED RESTRICTION DIGESTION MEDIATED CO-AMPLIFICATION

It is possible to use fingerprinting methods according to the invention that incorporate adapter ligation. Two or more populations being compared could be reverse transcribed with anchored oligodT primers possessing PCR™ primer binding sites and transcription promoters. The primer binding site would be the same for each of the primers but the transcription promoters would be different (T7 RNA polymerase for one, SP6 in another for instance). Following the tagging reaction, the populations could be mixed, restriction digested, ligated to an adapter, and amplified with primers specific to the adapter and the amplification domain of the tag. The amplified population would be split into multiple transcription reactions (*e.g.*, one with T7 RNA polymerase and one with SP6 RNA polymerase). The reactions would incorporate a labeled nucleotide and the resulting labeled nucleic acids would be fractionated (*e.g.*, on a sequencing polyacrylamide gel or an array). In addition to subtracting non-anchor-derived amplification products from the fingerprint population, the procedure would obviate the impact of tube-to-tube variation that is common in nucleic acid amplification reactions.

* * *

All of the compositions and/or methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

Beasley *et al.*, "Statistical Refinement of Primer Design Parameters," *In PCR Applications,* Innis, Gelfand, Sinsky (Eds.), pp. 55-72, 1999.

Butler and Chamberlin, *J. Biol. Chem.*, 257:5772-5778, 1982.

Chamberlin and Ryan, in The Enzymes, ed. P. Boyer Academic Press, New York, pp. 87-108, 1982.

Chapman, K. A. and Burgess, R. R., *Nucleic Acids Res.* 16:5413, 1987.

Chapman, K. A. and Wells, R. D., *Nucleic Acids Res.* 10(20):6331, 1982.

Chee *et al*, "Assessing genetic information with high density DNA Arrays," Science, 274:610-614, 1996.

Diaz *et al.*, *J. Mol. Biol.* 229: 805-811, 1993.

Dunn *et al.*, *Nature New Biology*, 230:94-96, 1971.

Dunn, J. J. and Studier, F. W., *J. Mol. Biol.* 166:477-535, 1983; and erratum *J. Mol. Biol.* 175:111-112, 1984.

Egholm *et al.*, "PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules," *Nature*, 365(6446):566-568, 1993.

European Patent Application No. 266,032

European Patent Application No. 329 822

European Patent Application No. 98302726

Froehler *et al.*, "Synthesis of DNA via deoxynucleoside H-phosphonate intermediates," *Nucleic Acids Res.* 14(13):5399-5407, 1986.

GB Application No. 2 202 328

Hausmann, *Current Topics in Microbiology and Immunology*, 75:77-109, 1976.

Ikeda *et al., Biochemistry* 32:9115-9124, 1993.

Innis *et al.*, "DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA," *Proc Natl Acad Sci USA.* 85(24):9436-9440, 1988.

Jones and Winistorfer, *Nuc. Acids. Res.* 20:595-600, 1992.

Kato, *Nuc Acids Res* 25:4694-4696, 1997.

Klement, *et al., J. Mol. Biol.* 215:21-29, 1990.

Kornberg, A. and T. A. Baker, DNA Replication, Second Edition, 1992, New York, W. H. Freeman and Company.

Korsten *et al.*, *J. Gen. Virol.*, 43:57-73, 1975.

Kotani *et al.* (1987), *Nucl. Acids Res.* 15:2653-2664, 1987.

Kwoh *et al.*, "Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format, *Proc Natl Acad Sci U S A.* 86(4):1173-1177, 1989.

Lizardi *et al.*, *Bio/technology* 6:1197-1202, 1988

Lomeli *et al.*, *Clin. Chem.* 35:1826-1831, 1989

Matoba, *et al.*, *Gene* 241:125-131, 2000.

Morris *et al.*, *Gene* 41:221-227, 1986.

PCT Application No. PCT/EP/01219

PCT Application No. PCT/US89/01025

PCT Application No. WO 88/10315

PCT Application No. WO 90/07641

PCT Application No. WO 92/20702

Sambrook *et al.*, *In: Molecular Cloning: A Laboratory Manual*, Vol. 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Ch. 7,7.19-17.29, 1989.

Scheit, *Nucleotide Analogs*, John Wiley, New York, 1980

Schneider, T. D. and Stormo, G. D., *Nucleic Acids Res.* 17(2):659, 1989.

Shuldiner *et al.*, Gene 91:139-142, 1990.

Towle *et al.*, *J. Biol. Chem.*, 250:1723-1733, 1975.

Trengle *et al.*, Nucl. Acids Res. 26:3883-3891, 1999.

Trenkle *et al.*, "Differential Display Probes for cDNA arrays," BioTechniques 27:554-564, 1999.

U.S. Patent 4,659,774

U.S. Patent 4,683,195

U.S. Patent 4,683,202

U.S. Patent 4,786,600

U.S. Patent 4,800,159

U.S. Patent 4,816,571

U.S. Patent 4,952,496

U.S. Patent 4,959,463

U.S. Patent 5,104,792
U.S. Patent 5,141,813
U.S. Patent 5,213,961
U.S. Patent 5,214,136
U.S. Patent 5,223,618
U.S. Patent 5,262,311
U.S. Patent 5,264,566
U.S. Patent 5,279,721
U.S. Patent 5,340,728
U.S. Patent 5,378,825
U.S. Patent 5,428,148
U.S. Patent 5,446,137
U.S. Patent 5,466,786
U.S. Patent 5,470,967
U.S. Patent 5,487,985
U.S. Patent 5,539,082
U.S. Patent 5,545,522
U.S. Patent 5,554,744
U.S. Patent 5,574,146
U.S. Patent 5,580,726
U.S. Patent 5,599,672
U.S. Patent 5,602,240
U.S. Patent 5,602,244
U.S. Patent 5,610,289
U.S. Patent 5,614,617
U.S. Patent 5,623,070
U.S. Patent 5,645,897
U.S. Patent 5,652,099
U.S. Patent 5,665,547
U.S. Patent 5,670,663
U.S. Patent 5,672,697
U.S. Patent 5,681,947
U.S. Patent 5,700,922

U.S. Patent 5,705,629
U.S. Patent 5,708,154
U.S. Patent 5,712,126
U.S. Patent 5,714,331
U.S. Patent 5,714,331
U.S. Patent 5,714,606
U.S. Patent 5,719,262
U.S. Patent 5,719,262
U.S. Patent 5,736,336
U.S. Patent 5,736,336
U.S. Patent 5,763,167
U.S. Patent 5,766,855
U.S. Patent 5,766,855
U.S. Patent 5,773,571
U.S. Patent 5,777,092
U.S. Patent 5,786,461
U.S. Patent 5,786,461
U.S. Patent 5,792,847
U.S. Patent 5,824,528
U.S. Patent 5,830,694
U.S. Patent 5,834,255
U.S. Patent 5,837,501
U.S. Patent 5,840,873
U.S. Patent 5,843,640
U.S. Patent 5,843,650
U.S. Patent 5,843,651
U.S. Patent 5,846,708
U.S. Patent 5,846,709
U.S. Patent 5,846,717
U.S. Patent 5,846,726
U.S. Patent 5,846,729
U.S. Patent 5,846,783
U.S. Patent 5,849,487

U.S. Patent 5,849,497
U.S. Patent 5,849,546
U.S. Patent 5,849,547
U.S. Patent 5,853,990
U.S. Patent 5,853,992
U.S. Patent 5,853,993
U.S. Patent 5,856,092
U.S. Patent 5,858,652
U.S. Patent 5,859,221
U.S. Patent 5,861,244
U.S. Patent 5,861,245
U.S. Patent 5,863,732
U.S. Patent 5,863,753
U.S. Patent 5,866,331
U.S. Patent 5,866,366
U.S. Patent 5,871,697
U.S. Patent 5,872,232
U.S. Patent 5,882,864
U.S. Patent 5,886,165
U.S. Patent 5,891,625
U.S. Patent 5,891,625
U.S. Patent 5,891,681
U.S. Patent 5,905,024
U.S. Patent 5,908,845
U.S. Patent 5,910,407
U.S. Patent 5,912,124
U.S. Patent 5,912,145
U.S. Patent 5,916,776
U.S. Patent 5,919,630
U.S. Patent 5,922,574
U.S. Patent 5,925,517
U.S. Patent 5,928,862
U.S. Patent 5,928,869

U.S. Patent 5,928,905

U.S. Patent 5,928,906

U.S. Patent 5,929,227

U.S. Patent 5,932,413

U.S. Patent 5,932,451

U.S. Patent 5,935,791

U.S. Patent 5,935,825

U.S. Patent 5,939,291

U.S. Patent 5,942,391

U.S. Patent 5,962,271

U.S. Patent 6,010,850

U.S. Patent 6,025,134

U.S. Patent 6,057,134

U.S. Patent 6,107,037

Watson *et al.*, Molecular Biology of The Gene, 4th Ed., Chapters 13-15, Benjamin/Cummings Publishing Co., Menlo Park, Calif.

Welsh and McClelland, *Nuc. Acids Res* 18:7213-7218, 1990.

CLAIMS

1. A method of analyzing one or more samples comprising:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a first nucleic acid tag, the nucleic acid tag comprising an amplification domain and a fingerprint domain;

c) amplifying the nucleic acid target using at least one adapter primer or arbitrary primer specific to a subset of the nucleic acid in the sample and one primer specific to the amplification domain, to produce at least a first amplified nucleic acid comprising the fingerprint domain and a segment of the first nucleic acid target;

d) generating labeled nucleic acid from the first amplified nucleic acid using the fingerprint domain; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

2. The method of claim 1, wherein the fingerprint domain of the first tag is appended between the at least a first nucleic acid target sequence and the amplification domain.

3. The method of claim 1, wherein the at least a first nucleic acid target is one of two or more nucleic acid targets.

4. The method of claim 1, wherein the at least a first sample is one of a plurality of samples.

5. The method of claim 1, wherein the first nucleic acid tag is appended to a target nucleic acid in each of the plurality of samples to prepare a plurality of tagged nucleic acid samples and each tagged nucleic acid sample is amplified in a separate reaction.

6. The method of claim 1, further defined as comprising:

a) obtaining at least a second sample;

b) preparing at least a second tagged nucleic acid sample by appending to at least the first nucleic acid target in the second sample at least a second nucleic acid tag, the second nucleic acid tag comprising the amplification

domain and a second fingerprint domain, wherein the second fingerprint domain differs from the first fingerprint domain;

c) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a mixture of tagged nucleic acid samples;

d) amplifying the first nucleic acid target in the mixture, using at least one arbitrary or adapter primer and one primer specific to the amplification domain, to produce mixed amplified nucleic acid comprising both amplified nucleic acid from the first nucleic acid sample that comprises the first fingerprint domain and a segment of the first nucleic acid target and amplified nucleic acid from the second nucleic acid sample that comprises the second fingerprint domain and a segment of the first nucleic acid target;

e) separating the mixed amplified nucleic acid into at least a first aliquot and a second aliquot;

f) generating first labeled nucleic acid from the first aliquot using the first fingerprint domain

g) generating second labeled nucleic acid from the second aliquot using the second fingerprint domain

h) fractionating the first labeled nucleic acid to create a first fingerprint; and

i) fractionating the second nucleic acid to create a second fingerprint.

7. The method of claim 6, further defined as comprising:

a) obtaining at least a third sample;

b) preparing at least a third tagged nucleic acid sample by appending to at least the first nucleic acid target in the third sample at least a third nucleic acid tag, the third nucleic acid tag comprising the amplification domain and a third fingerprint domain, wherein the third fingerprint domain differs from the first fingerprint domain and the second fingerprint domain;

c) mixing the first tagged nucleic acid sample, the second tagged nucleic acid sample, and the third tagged nucleic acid samples to create the mixture of tagged nucleic acid samples;

d) amplifying the first nucleic acid target in the mixture;

e) separating the mixed amplified nucleic acid into at least a first aliquot, a second aliquot, and a third aliquot;

f) generating first labeled nucleic acid, second labeled nucleic acid, and third labeled nucleic acid using the respective fingerprint domains; and

g) fractionating the first, second, and third, labeled nucleic acid to create respective first, second, and third fingerprints.

8. The method of claim 1 further defined as comprising:

a) obtaining a plurality of samples;

b) preparing a plurality of tagged nucleic acid samples by appending to at least the first nucleic acid target in each of the plurality of samples a plurality of nucleic acid tags, each nucleic acid tag comprising the amplification domain and a fingerprint domain, wherein the fingerprint domain for each sample is unique to that sample;

c) mixing the plurality of tagged nucleic acid samples to create a mixture of tagged nucleic acid samples;

d) amplifying the first nucleic acid target in the mixture, using at least one arbitrary primer or adaptor primer and one primer specific to the amplification domain, to produce amplified nucleic acid comprising a plurality of amplified nucleic acids, each of which comprises a fingerprint domain and a segment of the first nucleic acid target;

e) separating the amplified nucleic acid into a plurality of aliquots;

f) generating labeled nucleic acid from each aliquot using a fingerprint domain unique to a different sample for each aliquot; and

g) fractionating the labeled nucleic acid from each aliquot to create a plurality of fingerprints.

9. The method of claim 1, wherein the first nucleic acid target comprises RNA.

10. The method of claim 9, wherein the RNA comprises eukaryotic RNA or prokaryotic RNA.

11. The method of claim 1, wherein the first nucleic acid target comprises DNA.

12. The method of claim 1, wherein amplifying comprises a PCR™ amplification reaction.

13. The method of claim 1, wherein the primer specific to a subset of the nucleic acid in the sample is an arbitrary primer.

14. The method of claim 1, wherein the amplification domain comprises a primer binding domain.

15. The method of claim 14, wherein amplifying comprises a primer annealing to and being extended from the primer binding domain.

16. The method of claim 1, wherein at least a first nucleic acid adapter is appended to the first nucleic acid target.

17. The method of claim 16, wherein the first nucleic acid adapter comprises a primer binding domain.

18. The method of claim 17, wherein amplifying comprises a primer annealing to and being extended from the primer binding domain of the adapter.

19. The method of claim 1, wherein fractionating comprises electrophoretic separation of the labeled DNA.

20. The method of claim 1, wherein fractionating comprises array analysis of the labeled DNA.

21. The method of claim 1, wherein the fingerprint domain of the first nucleic acid tag or the second nucleic acid tag comprises a primer binding domain, a transcription domain, an affinity domain, or a combination thereof.

22. The method of claim 1, wherein the fingerprint domain comprises at least one primer binding domain.

23. The method of claim 22, wherein generating labeled nucleic acids comprises binding at least a first primer to at least one segment of the primer binding domain.

24. The method of claim 23, further comprising at least one primer extension reaction.

25. The method of claim 1, wherein the fingerprint domain comprises at least one transcription domain.

26. The method of claim 25, wherein the transcription domain comprises a promoter for a nucleic acid polymerase.

27. The method of claim 25, wherein generating labeled nucleic acids comprises at least one transcription reaction.

28. The method of claim 1, wherein the fingerprint domain comprises at least one affinity domain.

29. The method of claim 28, wherein generating labeled nucleic acids comprises binding at least a first ligand to at least a segment of the affinity domain.

30. The method of claim 29, wherein the first ligand comprises a nucleic acid.

31. The method of claim 29, wherein the first ligand is bound to at least one solid support.

32. The method of claim 29, wherein the first ligand is labeled.

33. The method of claim 32, wherein the binding of the first ligand to the segment of the affinity domain results in a detectable signal.

34. The method of claim 1, wherein the first nucleic acid tag further comprises at least one additional domain.

35. The method of claim 34, wherein the additional domain is a labeling domain, a restriction enzyme domain, a secondary amplification domain, a secondary fingerprint domain or a combination thereof.

36. The method of claim 34, wherein the additional domain comprises at least one labeling domain.

37. The method of claim 36, wherein the labeling domain is comprised between the fingerprint domain and the amplification domain.

38. A method of analyzing two or more samples comprising:

a) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of at least a first sample at least a first nucleic acid tag, the first nucleic acid tag comprising an amplification domain and a first fingerprint domain;

b) preparing at least a second tagged nucleic acid sample by appending to at least the first nucleic acid target in at least a second sample at least a second nucleic acid tag, the second nucleic acid tag comprising the amplification domain and a second fingerprint domain, wherein the second fingerprint domain differs from the first fingerprint domain;

c) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a mixture of tagged nucleic acid samples;

d) amplifying the first nucleic acid target in the mixture, using at least one arbitrary primer and one primer specific to the amplification domain, to produce amplified nucleic acid comprising both amplified nucleic acid from the first nucleic acid sample that comprises the first fingerprint domain and a segment of the first nucleic acid target and amplified nucleic acid from the second nucleic acid sample that comprises the second fingerprint domain and a segment of the first nucleic acid target in the second sample;

e) separating the amplified nucleic acid into at least a first aliquot and a second aliquot;

f) generating first labeled nucleic acid from the first aliquot using the first fingerprint domain

g) generating second labeled nucleic acid from the second aliquot using the second fingerprint domain

h) fractionating the first labeled nucleic acid to create a first fingerprint; and

i) fractionating the second nucleic acid to create a second fingerprint.

39. A method of analyzing one or more samples comprising:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a nucleic acid tag, the nucleic acid tag comprising an amplification domain and a fingerprint domain, wherein the fingerprint domain is appended between the nucleic acid target sequence and the amplification domain;

c) amplifying the nucleic acid target using at least one arbitrary primer and one primer specific to the amplification domain, to produce at least a first amplified nucleic acid comprising the fingerprint domain and a segment of the first nucleic acid target;

d) generating labeled nucleic acid from the first amplified nucleic acid using the fingerprint domain; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

40. A method of analyzing one or more samples comprising:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a first nucleic acid tag, the first nucleic acid tag comprising an amplification domain and a first fingerprint domain, wherein the fingerprint domain comprises a primer binding site and is appended between the nucleic acid target sequence and the amplification domain;

c) amplifying the nucleic acid target using at least one adaptor primer or arbitrary primer and one primer specific to the amplification domain, to

produce at least a first amplified nucleic acid comprising the first fingerprint domain and a segment of the first nucleic acid target;

d) generating labeled nucleic acid from the first amplified nucleic acid by primer extension from the primer binding site; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

41. A method of analyzing one or more samples comprising:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a first nucleic acid tag, the first nucleic acid tag comprising an amplification domain and a first fingerprint domain, wherein the fingerprint domain comprises a transcription domain and is appended between the nucleic acid target sequence and the amplification domain;

c) amplifying the nucleic acid target using at least one adaptor primer or arbitrary primer and one primer specific to the amplification domain, to produce at least a first amplified nucleic acid comprising the first fingerprint domain and a segment of the first nucleic acid target;

d) generating labeled nucleic acid from the first amplified nucleic acid using a nucleic acid polymerase specific to the transcription domain; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

42. A method of analyzing one or more samples comprising:

a) obtaining a first sample;

b) preparing at least a first tagged nucleic acid sample by appending to at least a first nucleic acid target of the first sample at least a first nucleic acid tag, the first nucleic acid tag comprising an amplification domain and a first fingerprint domain, wherein the fingerprint domain comprises an affinity domain and is appended between the nucleic acid target sequence and the amplification domain;

c) amplifying the nucleic acid target using at least one adaptor primer or arbitrary primer and one primer specific to the amplification domain, to produce at least a first amplified nucleic acid comprising the first fingerprint domain and a segment of the first nucleic acid target;

d) isolating or labeling nucleic acid from the first amplified nucleic acid using the affinity domain; and

e) fractionating the labeled nucleic acid to create a fingerprint of the sample.

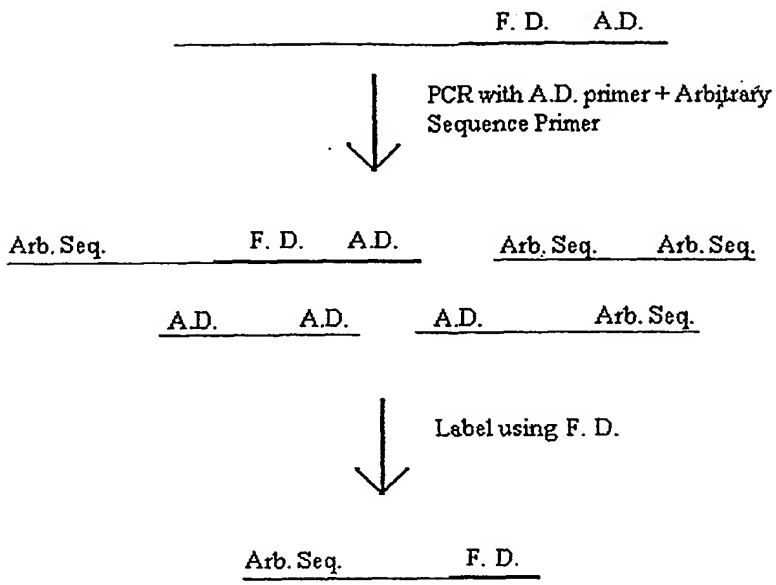
F. D. A.D.
PCR with A.D. primer + Arbitrary Sequence Primer
Arb. Seq. F. D. A.D.
Arb. Seq. Arb. Seq.
A.D. A.D.
A.D. Arb. Seq.
Label using F. D.
Arb. Seq. F. D.

FIG. 1

Nucleic acid population	F. D.	A. D.

FIG. 2

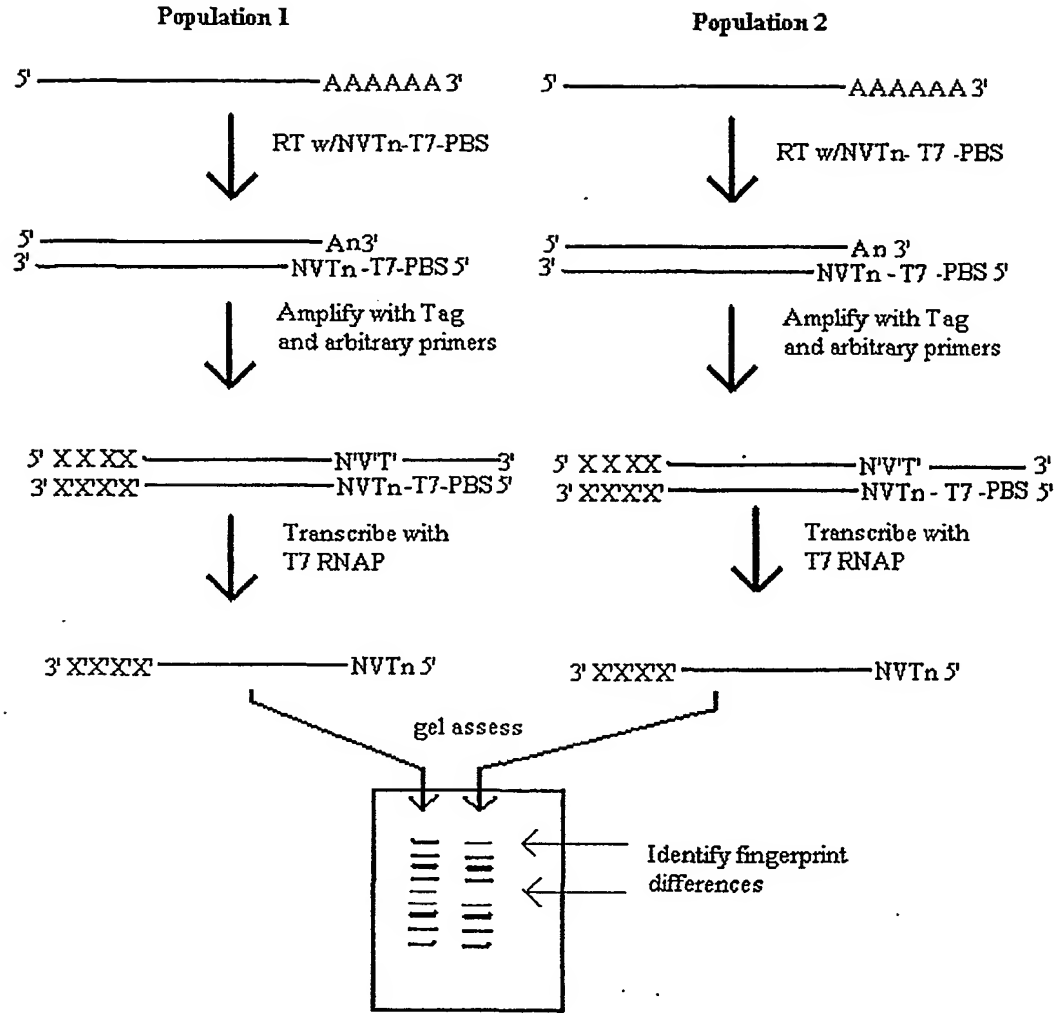
Population 1
Population 2
5' AAAAAA 3'
RT w/NVTn-T7-PBS
RT w/NVTn- T7 -PBS
An3'
NVTn-T7-PBS 5'
An 3'
NVTn - T7 -PBS 5'
Amplify with Tag and arbitrary primers
5' XXXX N'V'T' 3'
3' X'X'X'X' NVTn-T7-PBS 5'
NVTn - T7 -PBS 5'
Transcribe with T7 RNAP
3' X'X'X'X' NVTn 5'
gel assess
Identify fingerprint differences

FIG. 3

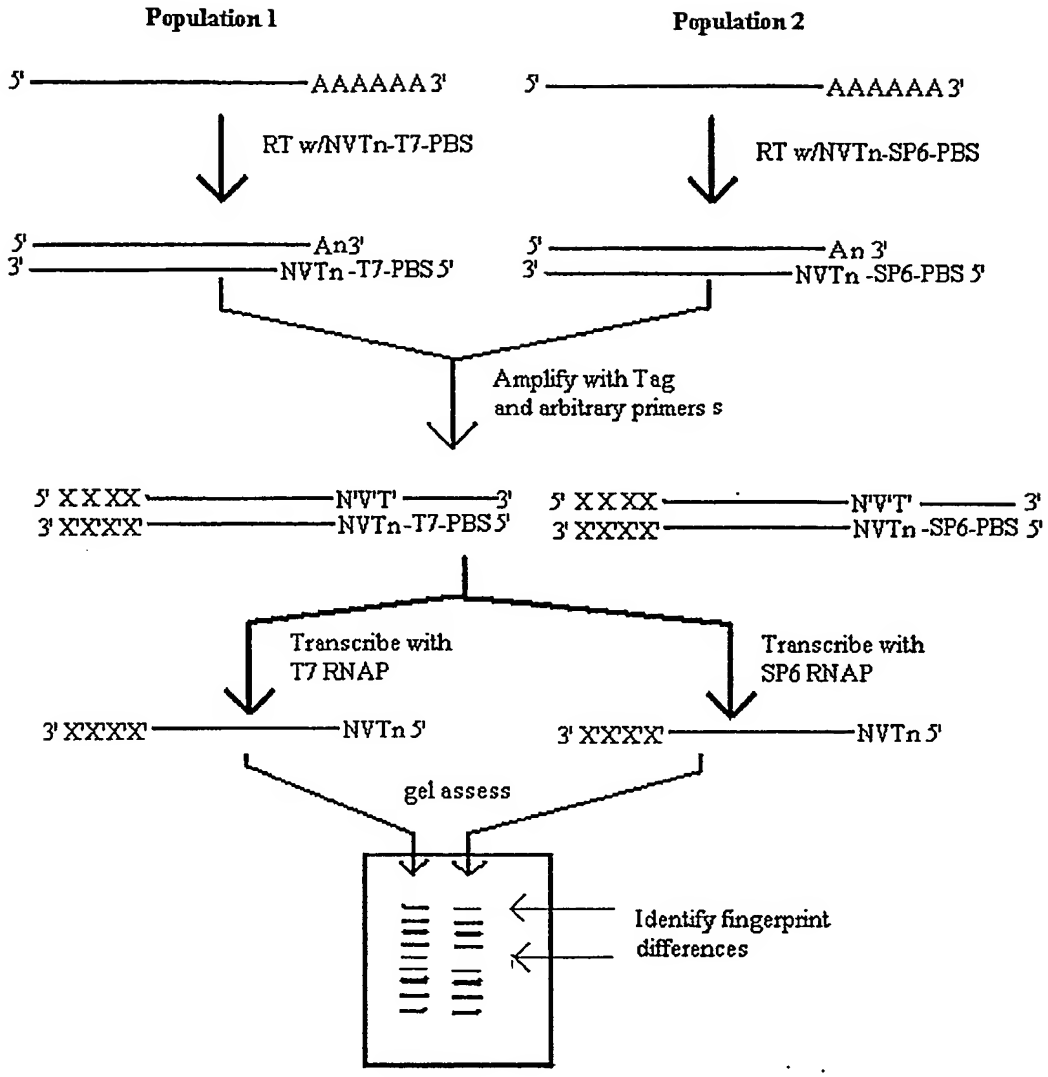
Population 1
Population 2
5' AAAAAA 3'
5' AAAAAA 3'
RT w/NVTn-T7-PBS
RT w/NVTn-SP6-PBS
5' An3'
3' NVTn -T7-PBS 5'
5' An 3'
3' NVTn -SP6-PBS 5'
Amplify with Tag
and arbitrary primers s
5' XXXX N'V'T' 3'
3' X'X'X'X' NVTn-T7-PBS 5'
5' XXXX N'V'T' 3'
3' X'X'X'X' NVTn -SP6-PBS 5'
Transcribe with
T7 RNAP
Transcribe with
SP6 RNAP
3' X'X'X'X' NVTn 5'
3' X'X'X'X' NVTn 5'
gel assess
Identify fingerprint
differences

FIG. 4

SEQUENCE LISTING

<110> BROWN, DAVID
WINKLER, MATTHEW M.
LAWRENCE, FRANK
SHELTON, JEFFREY

<120> METHODS FOR GENERATING NUCLEIC ACID FINGERPRINTS

<130> AMBI:058-WO

<140> 60/265,693
<141> 2001-01-31

<160> 12

<170> PatentIn Ver. 2.1

<210> 1
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<221> modified_base
<222> (60)
<223> V = G OR C OR A

<220>
<221> modified_base
<222> (61)
<223> N = A OR C OR G OR T/U

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 1
gctgatggcg atgaatgaac actgtaatac gactcactat agggagattt ttttttttv 60
n 61

<210> 2
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<221> modified_base
<222> (60)
<223> V = G OR C OR A

<220>
<221> modified_base
<222> (61)
<223> N = A OR C OR G OR T/U

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 2
gctgatggcg atgaatgaac actgatttag gtgacactat agaagagttt ttttttttv 60
n 61

<210> 3
<211> 168
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 3
taatacgact cactataggg ttcgggctta ggctccagtg cctgttcggt ggtcgcggcg 60
ctgatggcga tgaatgaaca ctgcggcaag ccgcttaatg acactcgttt gctggctttg 120
atgggcgagc tggaaggccg tatctccggc agcattcatt acgacaaa 168

<210> 4
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 4
gctgatggcg atgaatgaac actg 24

<210> 5
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 5
agcagttggt tggagcaaca tc 22

<210> 6
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 6
gacactcgtt tgctggcttt gatg 24

<210> 7
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 7
gctgatggcg atgaatgaac actg 24

<210> 8
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 8
cagggacctt ctctaagact ttc 23

<210> 9
<211> 68
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 9
gctgatggcg atgaatgaac actgtaatac gactcactat agggagaaca tgggtaatcc 60
tcatgttt 68

<210> 10
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 10
gctgatggcg atgaatgaac a 21

<210> 11
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

Primer

<400> 11
cctcagcaat cgcagcaaac t 21

<210> 12
<211> 68
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 12
gctgatggcg atgaatgaac actgatttag gtgacactat agaagacaca tgggtaatcc 60
tcatgttt 68